

Fakultät für Wirtschaftswissenschaften und Medien

Institut für Betriebswirtschaftslehre

Fachgebiet Steuerlehre/Prüfungswesen

Dissertation

Benford's Law – Eine numerische Untersuchung zur Anwendbarkeit statistischer Anpassungstests und praktische Anwendung auf volkswirtschaftliche Zahlen Deutschlands sowie Kfz-Versicherungsschäden

verfasst von: Daniel Brinkert, M.Sc.

Erstgutachter: Herr Univ.-Prof. Dr. habil. Gernot Brähler

Zweitgutachter: Herr Univ.-Prof. Dr. habil. Michael Grüning

Datum der Einreichung: 28.02.2016

Eröffnung des Promotionsverfahrens am: 03.05.2016

Datum der wissenschaftlichen Aussprache: 30.11.2016

URN: urn:nbn:de:gbv:ilm1-2016000669

Inhaltsverzeichnis

Abbildungsverzeichnis	IV
Tabellenverzeichnis	VI
Symbolverzeichnis	XII
Abkürzungsverzeichnis	XV
Vorwort.....	XVIII
1. Benford's Law – Eine numerische Untersuchung zur Anwendbarkeit statistischer Anpassungstests und praktische Anwendung auf volkswirtschaftliche Zahlen Deutschlands sowie Kfz-Versicherungsschäden.....	1
1.1 Einleitung	1
1.2 Problemstellung.....	2
1.3 Zielstellung	7
1.4 Gang der Arbeit	9
2. Aktueller Stand der Forschung zu Benford's Law	11
2.1 Anwendung in der Rechnungslegung sowie bei Steuererklärungen und –prüfungen.	11
2.2 Analyse von makroökonomischen und statistischen Daten	15
2.3 Sonstige Anwendungen	17
3. Grundlagen des Benford's Law	21
3.1 Beschreibung der Gesetzesmäßigkeit.....	21
3.2 Voraussetzungen und Eigenschaften.....	23
3.3 Eignung als Indikator von Manipulationen	25
3.4 Statistische Testverfahren und Maßzahlen	28
3.4.1 Der χ^2 -Anpassungstest	29
3.4.2 Kolmogorow-Smirnow-Anpassungstest	34
3.4.3 Kuiper-Test	36
3.4.4 Test der Z-Statistik.....	38
3.4.5 Normierter χ^2 -Wert.....	39

3.4.6 Normierter euklidischer Abstand	40
3.4.7 Abstand über die arithmetischen Mittelwerte	42
3.4.8 Manhattan-Distanz und Mean-Absolute-Deviation	43
3.4.9 Tschebyscheff-Distanz	45
3.4.10 Verzerrungsfaktor-Modell	46
4. Prüfung der Eignungsqualität der Anpassungstests und statistischen Kennzahlen zur Bestimmung der Abweichungen von Benford's Law	48
4.1 Beschreibung der Methodik und Vorgehensweise	48
4.2 Die Monte-Carlo-Simulation	49
4.3 Vergleich der Resultate der einzelnen Verfahren	58
4.3.1 Ergebnisse des χ^2 -Anpassungstestes	63
4.3.2 Ergebnisse des KS-Tests	67
4.3.3 Ergebnisse des Kuiper-Tests	70
4.3.4 Ergebnisse der Z-Statistik	72
4.3.5 Ergebnisse des normierten χ^2 -Wertes	77
4.3.6 Ergebnisse des normierten euklidischen Abstandes d^*	79
4.3.7 Ergebnisse des Abstands über die arithmetischen Mittelwerte a^*	81
4.3.8 Ergebnisse der Manhattan-Distanz	84
4.3.9 Ergebnisse der Tschebyscheff-Distanz	86
4.3.10 Ergebnisse des Verzerrungsfaktor-Modells	88
4.4 Auswertung und Schlussfolgerungen der Untersuchung	89
5. Vergleich der Qualität historischer makroökonomischer Daten der Bundesrepublik Deutschland und ihrer Vorgängerstaaten	95
5.1 Grundlagen des Datenmaterials	95
5.2 Deskriptive Statistik und Untersuchungsdesign	96
5.3 Ergebnisse der Untersuchung	101
5.3.1 Ergebnisse des aggregierten Datensatzes	101

5.3.2 Ergebnisse der einzelnen Staatsformen.....	103
5.3.3 Ergebnisse der Einzeljahre.....	107
5.4 Bewertung der Ergebnisse.....	113
6. Anwendung des Benford's Law zur Manipulationsaufdeckung bei Kfz- Versicherungsschäden	120
6.1 Grundlagen des Datenmaterials.....	120
6.2 Deskriptive Statistik und Untersuchungsdesign.....	121
6.3 Ergebnisse der Untersuchung	125
6.3.1 Ergebnisse der Gesamtdatensätze	125
6.3.2 Ergebnisse des Datensatzes der Partnerwerkstätten	129
6.3.3 Ergebnisse des Datensatzes des Expertisechecks	145
6.4 Bewertung der Ergebnisse.....	151
7. Zusammenfassung und Ausblick.....	156
Anhang	164
Literaturverzeichnis.....	XIX
Rechtsprechungsverzeichnis.....	XXVII
Quellenverzeichnis	XXVIII
Eidesstattliche Erklärung.....	XXXI

Abbildungsverzeichnis

Abbildung 1: Verteilung der ersten Ziffer d_1 des Benfordschen Gesetzes.....	22
Abbildung 2: Vergleich der Verteilung der ersten Ziffern von Hill, Benford und der Gleichverteilung	26
Abbildung 3: Berechnung des Abstandes zweier Punkte mittels des Satzes des Pythagoras	41
Abbildung 4: Grafische Illustration der Manhattan-Distanz zwischen zweier Punkte	44
Abbildung 5: Dichtefunktion der Standardnormalverteilung mit Erwartungswert von null und Standardabweichung von eins	56
Abbildung 6: Häufigkeitsverteilung der mittels Monte-Carlo-Simulation erzeugten Zufallszahlen für $n = 500$	56
Abbildung 7: Häufigkeitsverteilung der mittels Monte-Carlo-Simulation erzeugten Zufallszahlen für $n = 10.000$	57
Abbildung 8: Häufigkeitsverteilung der mittels Monte-Carlo-Simulation erzeugten Zufallszahlen für $n=1.000.000$	58
Abbildung 9: Verteilung der ersten Ziffern der Benford-verteilten Datensätze.....	60
Abbildung 10: Verteilung der zweiten Ziffern der Benford-verteilten Datensätze.....	61
Abbildung 11: Häufigkeitsverteilung der ersten beiden Ziffern des Benford-verteilten Datensatzes mit $n = 150$ Daten.....	62
Abbildung 12: Häufigkeitsverteilung der ersten beiden Ziffern des Benford-verteilten Datensatzes mit $n = 5.000$ Daten.....	62
Abbildung 13: Häufigkeitsverteilung der ersten beiden Ziffern des Benford-verteilten Datensatzes mit $n = 1.000.000$ Daten.....	63
Abbildung 14: Verteilung der ersten Ziffern des Gesamtdatensatzes der historischen Daten.....	101
Abbildung 15: Verteilung der zweiten Ziffern des Gesamtdatensatzes der historischen Daten.....	102
Abbildung 16: Ziffernverteilung der ersten Ziffern der einzelnen Staatsformen.....	104
Abbildung 17: Ziffernverteilung der zweiten Ziffern der einzelnen Staatsformen	105
Abbildung 18: Verteilung der ersten Ziffern der auffälligen Jahre der Weimarer Republik	108
Abbildung 19: Verteilung der zweiten Ziffern der auffälligen Jahre der Weimarer Republik	108

Abbildung 20: Verteilung der ersten Ziffern der auffälligen Jahre des NS-Regimes	110
Abbildung 21: Verteilung der zweiten Ziffern der auffälligen Jahre des NS-Regimes	110
Abbildung 22: Verteilung der ersten Ziffern der Gesamtdatensätze	126
Abbildung 23: Verteilung der zweiten Ziffern der Gesamtdatensätze	127
Abbildung 24: Verteilung der zweiten Ziffern der freien Werkstätten der Jahre 2013 und 2014	131
Abbildung 25: Verteilung der zweiten Ziffern von Premiumhersteller C und Massenhersteller A in den Jahren 2013 und 2014	134
Abbildung 26: Verteilung der zweiten Ziffern des Datensatzes Expertisecheck des Jahres 2013 bei Unterscheidung nach Stand der Prüfung	148
Abbildung 27: Verteilung der zweiten Ziffern des Datensatzes Expertisecheck des Jahres 2014 bei Unterscheidung nach Stand der Prüfung	149

Tabellenverzeichnis

Tabelle 1: Wahrscheinlichkeiten des Auftretens der ersten bis vierten Ziffer der Benford-Verteilung	23
Tabelle 2: Näherungsformeln für den kritischen Wert des KS-Tests bei stetigen Verteilungen sowie der Benford-Verteilung	35
Tabelle 3: Näherungsformeln für den kritischen Wert des Kuiper-Tests bei stetigen Verteilungen sowie der Benford-Verteilung	37
Tabelle 4: Kritische Werte der Z-Statistik.....	38
Tabelle 5: Richtwerte zur Beurteilung der MAD	45
Tabelle 6: Statistische Kennwerte der generierten Benford-verteilten Datensätze	59
Tabelle 7: Ergebnisse des χ^2 -Anpassungstestes bei Betrachtung der ersten Ziffern	65
Tabelle 8: Ergebnisse des χ^2 -Anpassungstestes bei Betrachtung der zweiten Ziffer	66
Tabelle 9: Ergebnisse des χ^2 -Anpassungstestes bei Betrachtung der ersten beiden Ziffern.....	66
Tabelle 10: kritische Werte des KS-Testes in Abhängigkeit der Stichprobenumfänge	67
Tabelle 11: Ergebnisse des KS-Testes bei Betrachtung der ersten Ziffern	68
Tabelle 12: Ergebnisse des KS-Testes bei Betrachtung der zweiten Ziffern	69
Tabelle 13: Ergebnisse des KS-Tests bei Betrachtung der beiden Ziffern.....	69
Tabelle 14: Ergebnisse des Kuiper-Testes bei Betrachtung der ersten Ziffern	71
Tabelle 15: Ergebnisse des Kuiper-Testes bei Betrachtung der zweiten Ziffern	71
Tabelle 16: Ergebnisse des Kuiper-Testes bei Betrachtung der ersten beiden Ziffern	72
Tabelle 17: Ergebnisse der Z-Statistik bei Analyse der ersten Ziffern.....	74
Tabelle 18: Ergebnisse der Z-Statistik bei Analyse der zweiten Ziffern.....	75
Tabelle 19: Ergebnisse der Z-Statistik bei Analyse der ersten beiden Ziffern.....	76
Tabelle 20: Ergebnisse des normierten χ^2 -Wertes bei Analyse der ersten Ziffern	78
Tabelle 21: Ergebnisse des normierten χ^2 -Wertes bei Analyse der zweiten Ziffern	78
Tabelle 22: Ergebnisse des normierten χ^2 -Wertes bei Analyse der ersten beiden Ziffern	79
Tabelle 23: Ergebnisse des normierten euklidischen Abstandes bei Analyse der ersten Ziffern.....	80
Tabelle 24: Ergebnisse des normierten euklidischen Abstandes bei Analyse der zweiten Ziffern.....	80

Tabelle 25: Ergebnisse des normierten euklidischen Abstandes bei Analyse der ersten beiden Ziffern	81
Tabelle 26: Ergebnisse des Abstands über die arithmetischen Mittelwerte bei Betrachtung der ersten Ziffern	82
Tabelle 27: Ergebnisse des Abstands über die arithmetischen Mittelwerte bei Betrachtung der zweiten Ziffern	83
Tabelle 28: Ergebnisse des Abstands über die arithmetischen Mittelwerte bei Betrachtung der ersten beiden Ziffern.....	83
Tabelle 29: Ergebnisse der Manhattan-Distanz bei Analyse der ersten Ziffern.....	85
Tabelle 30: Ergebnisse der Manhattan-Distanz bei Analyse der zweiten Ziffern.....	85
Tabelle 31: Ergebnisse der Manhattan-Distanz bei Analyse der ersten beiden Ziffern	86
Tabelle 32: Ergebnisse der Tschebyscheff-Distanz bei Analyse der ersten Ziffern.....	87
Tabelle 33: Ergebnisse der Tschebyscheff-Distanz bei Analyse der zweiten Ziffern.....	87
Tabelle 34: Ergebnisse der Tschebyscheff-Distanz bei Analyse der ersten beiden Ziffern	88
Tabelle 35: Ergebnisse des Verzerrungsfaktor-Modells	89
Tabelle 36: Generelle Eignung der Testgröße zur Applikation im Rahmen des Benford-Testes.....	91
Tabelle 37: Praktische Erfüllung der Eigenschaft der Stichprobeninvarianz einzelner Testverfahren.....	92
Tabelle 38: Auf Basis der Untersuchung empfohlene Größe des zu untersuchenden Stichprobenumfanges der einzelnen Testgrößen.....	93
Tabelle 39: Stichprobenumfang und statistische Kennzahlen des Gesamtdatensatzes und der einzelnen Staatsformen.....	99
Tabelle 40: Ergebnisse des Gesamtdatensatzes der ersten und zweiten Ziffern	102
Tabelle 41: Rangzahlen der Testgrößen der unterschiedlichen Staatsformen bei Analyse der ersten Ziffern	103
Tabelle 42: Ergebnisse der Z-Statistik der Staatsformen bei Analyse der ersten Ziffern	104
Tabelle 43: Rangzahlen der Testgrößen der unterschiedlichen Staatsformen bei Analyse der ersten Ziffern	105
Tabelle 44: Ergebnisse der Z-Statistik der Staatsformen bei Analyse der zweiten Ziffern ...	106
Tabelle 45: Ergebnisse des Verzerrungsfaktor-Modells der Staatsformen	106
Tabelle 46: Stichprobenumfang und statistische Kennzahlen der untersuchten Datensätze der Kfz-Unfallschäden.....	123

Tabelle 47: Ergebnisse der Z-Statistik für die ersten Ziffern der Gesamtdatensätze	126
Tabelle 48: Ergebnisse der statistischen Testgrößen der Gesamtdatensätze für die ersten Ziffern.....	128
Tabelle 49: Ergebnisse der statistischen Testgrößen der Gesamtdatensätze für die zweiten Ziffern.....	128
Tabelle 50: Rangzahlen der Testergebnisse der zweiten Ziffern des Datensatzes der Partnerwerkstätten bei Unterscheidung nach Art der Werkstatt	129
Tabelle 51: Ergebnisse des Verzerrungsfaktor-Modells bei Unterscheidung nach Art der Werkstatt.....	130
Tabelle 52: Ergebnisse der Z-Statistik der zweiten Ziffern bei Unterscheidung nach Art der Werkstatt.....	130
Tabelle 53: Ergebnisse des Benford-Testes in Form von Rangzahlen bei Applikation auf die zweiten Ziffern bei Unterscheidung nach Kfz-Hersteller.....	132
Tabelle 54: Ergebnisse des Verzerrungs-Faktormodells bei Unterscheidung nach Kfz- Hersteller	133
Tabelle 55: Ergebnisse der Z-Statistik der zweiten Ziffern der Hersteller Premiumhersteller C und Massenhersteller A	134
Tabelle 56: Ergebnisse des Benford-Testes in Form von Rangzahlen bei Applikation auf die zweiten Ziffern der freien Werkstätten bei Unterscheidung nach Kfz-Hersteller.....	135
Tabelle 57: Ergebnisse des Verzerrungs-Faktormodells der freien Werkstätten bei Unterscheidung nach Kfz-Hersteller	136
Tabelle 58: Ergebnisse des Benford-Testes in Form von Rangzahlen bei Applikation auf die zweiten Ziffern der Markenwerkstätten bei Unterscheidung nach Kfz-Hersteller.....	137
Tabelle 59: Ergebnisse des Verzerrungs-Faktormodells der Markenwerkstätten bei Unterscheidung nach Kfz-Hersteller	138
Tabelle 60: Ergebnisse der Z-Statistik für die zweiten Ziffern der Markenwerkstätten von Premiumhersteller C	138
Tabelle 61: Ergebnisse des Benford-Testes in Form von Rangzahlen bei Applikation auf die zweiten Ziffern bei Unterscheidung nach Art des Schadens	139
Tabelle 62: Ergebnisse des Verzerrungs-Faktormodells bei Unterscheidung nach Art des Schadens	140
Tabelle 63: Ergebnisse des Benford-Testes in Form von Rangzahlen bei Applikation auf die zweiten Ziffern der freien Werkstätten bei Unterscheidung nach Art des Schadens	141

Tabelle 64: Ergebnisse des Verzerrungs-Faktormodells der freien Werkstätten bei Unterscheidung nach Art des Schadens.....	141
Tabelle 65: Ergebnisse des Benford-Testes in Form von Rangzahlen bei Applikation auf die zweiten Ziffern der Markenwerkstätten bei Unterscheidung nach Art des Schadens	142
Tabelle 66: Ergebnisse des Verzerrungs-Faktormodells der Markenwerkstätten bei Unterscheidung nach Art des Schadens.....	142
Tabelle 67: Ergebnisse der zweiten Ziffern in Form von Rangzahlen des Datensatzes der Partnerwerkstätten bei Unterscheidung nach Datum der Erstzulassung des Unfall-Kfz	144
Tabelle 68: Ergebnisse des Verzerrungs-Faktormodells bei Unterscheidung nach Datum der Erstzulassung	145
Tabelle 69: Ergebnisse der zweiten Ziffern in Form von Rangzahlen des Datensatzes „Expertisecheck“ bei Unterscheidung nach Stand der Prüfung	146
Tabelle 70: Ergebnisse des Verzerrungsfaktor-Modells des Datensatzes „Expertisecheck“ bei Unterscheidung nach Stand der Prüfung	147
Tabelle 71: Ergebnisse der Z-Statistik der zweiten Ziffern des Datensatzes des Expertisechecks	147
Tabelle 72: Ergebnisse der zweiten Ziffern des Datensatzes des Expertisechecks bei Unterteilung nach Datum der Kfz-Erstzulassung	150
Tabelle 73: Ergebnisse des Verzerrungsfaktor-Modells des Datensatzes „Expertisecheck“ bei Unterscheidung nach Datum der Erstzulassung des Unfall-Kfz	151
Tabelle 74: Zusammenfassung der Ergebnisse der Anwendung des Benford-Testes auf die zweiten Ziffern der Kfz-Schadensfälle	154
Tabelle 75: Statistische Kennzahlen und Stichprobenumfänge der Einzeljahre der Weimarer Republik:	166
Tabelle 76: Statistische Kennzahlen und Stichprobenumfänge der Einzeljahre des NS-Regimes:	167
Tabelle 77: Statistische Kennzahlen und Stichprobenumfänge der Einzeljahre der Deutschen Demokratischen Republik:	167
Tabelle 78: Statistische Kennzahlen und Stichprobenumfänge der Einzeljahre der Bundesrepublik Deutschland:	169
Tabelle 79: Ergebnisse des Benford-Testes der einzelnen Staatsformen bei Analyse der ersten Ziffern:	171

Tabelle 80: Ergebnisse des Benford-Testes der einzelnen Staatsformen bei Analyse der zweiten Ziffern:	171
Tabelle 81: Ergebnisse des Benford-Testes bei Anwendung auf die ersten Ziffern der Weimarer Republik (Jahre mit größter Abweichung fett markiert):	172
Tabelle 82: Ergebnisse des Benford-Testes bei Anwendung auf die zweiten Ziffern der Weimarer Republik (Jahre mit größter Abweichung fett markiert):	172
Tabelle 83: Ergebnisse des Benford-Testes bei Anwendung auf die ersten Ziffern des NS-Regimes (Jahre mit größter Abweichung fett markiert):	173
Tabelle 84: Ergebnisse des Benford-Testes bei Anwendung auf die zweiten Ziffern des NS-Regimes (Jahre mit größter Abweichung fett markiert):	173
Tabelle 85: Vergleich der Ergebnisse des Benford-Testes zwischen der DDR und der BRD bei Anwendung auf die ersten Ziffern (höhere Abweichung im Einzeljahr fett markiert):	174
Tabelle 86: Vergleich der Ergebnisse des Benford-Testes zwischen der DDR und der BRD bei Anwendung auf die zweiten Ziffern (höhere Abweichung im Einzeljahr fett markiert):	177
Tabelle 87: Ergebnisse des Benford-Testes der zweiten Ziffern bei Unterscheidung nach Art der Werkstatt:	180
Tabelle 88: Ergebnisse der Analyse der zweiten Ziffern bei Unterscheidung nach Kfz-Hersteller	181
Tabelle 89: Ergebnisse der Analyse der zweiten Ziffern der freien Werkstätten bei Unterscheidung nach Kfz-Hersteller:	182
Tabelle 90: Ergebnisse der Analyse der zweiten Ziffern der Markenwerkstätten bei Unterscheidung nach Kfz-Hersteller:	183
Tabelle 91: Ergebnisse des Benford-Testes bei Anwendung auf die zweiten Ziffern bei Unterscheidung nach Art des Schadens:	183
Tabelle 92: Ergebnisse des Benford-Testes bei Anwendung auf die zweiten Ziffern der freien Werkstätten bei Unterscheidung nach Art des Schadens:	184
Tabelle 93: Ergebnisse des Benford-Testes bei Anwendung auf die zweiten Ziffern der Markenwerkstätten bei Unterscheidung nach Art des Schadens:	184
Tabelle 94: Ergebnisse des Benford-Testes bei Anwendung auf die zweiten Ziffern bei Unterscheidung nach Datum der Erstzulassung	185

Tabelle 95: Ergebnisse der zweiten Ziffern des Datensatzes „Expertisecheck“ bei Unterscheidung nach Stand der Prüfung	186
Tabelle 96: Ergebnisse des Benford-Testes bei Anwendung auf die zweiten Ziffern des Datensatzes „Expertisecheck“ bei Unterscheidung nach Datum der Erstzulassung	187

Symbolverzeichnis

€	Euro (Währungseinheit)
%	Prozent
§	Paragraph
§§	Paragraphen
n	Stichprobenumfang
d_n	Ziffer an der n -ten Stelle (empirische Größe in Stichprobe)
D_n	Ziffer an der n -ten Stelle (Zufallsvariable)
$P(D)$	Wahrscheinlichkeit abhängig von D
\log_{10}	Logarithmus zur Basis 10
\in	Element
\mathbb{Z}	ganze Zahlen
\mathbb{R}	reelle Zahlen
$\sum x_i$	Summe über x_i
α	Signifikanzniveau
T	Testgröße
χ^2	Chi-Quadrat (Prüfgröße)
H_0	Nullhypothese
H_1	Gegenhypothese
P_0	Theoretische Verteilung, angenommen bei Nullhypothese
K^*	Ablehnungsbereich
$\chi^2_{m-1,r;1-\alpha}$	Untere Grenze des Ablehnungsbereiches des χ^2 -Testes; Abhängig von m , r und α
m	Anzahl der Merkmalsklassen
r	Anzahl geschätzter Parameter
f	Freiheitsgrade
∞	unendlich
h_j	absolute Häufigkeit der Klasse j
f_j	relative Häufigkeit der Klasse j
p_j	hypothetische Klassenwahrscheinlichkeit der Klasse j
c	χ^2 -Wert dividiert durch Stichprobenumfang
d^*	normierte euklidische Abstandsmessung

a^*	Abstandsmessung der arithmetischen Mittelwerte
μ_e	arithmetischer Mittelwert der empirischen Beobachtungswerte
μ_b	arithmetischer Mittelwert der Klassenmitten der unterstellten Verteilung
E_j	Erwartungshäufigkeit der Klasse j
k	Maximale Stellenanzahl einer Zahl
i	Stellenzähler
D	Testgröße des KS-Testes
$F_0(x)$	unterstellte Verteilungsfunktion
$F_e(x)$	empirische Verteilungsfunktion
$S_i(x)$	Relative Summenfunktion
\max_x	Maximum über x
V_N	Testgröße des Kuiper-Testes
V_N^*	Modifizierte Testgröße des Kuiper-Testes
Z	Testgröße der Z-Statistik
d_{ij}	Abstand zwischen dem Punkt i und j
x_i	Punkt i
MD	Manhattan-Distanz
MAD	Mean-Absolute-Deviation (mittlere absolute Abweichung)
TD	Tschebyscheff-Distanz
VF	Verzerrungsfaktor
TM	tatsächliche Mittelwert
EM	erwarteter Mittelwert
Ω	Ergebnismenge im Wahrscheinlichkeitsraum
Ω'	Ergebnismenge im Messraum
Σ	Potenzmenge im Wahrscheinlichkeitsraum
Σ'	Potenzmenge im Messraum
P	Wahrscheinlichkeitsmaß im Wahrscheinlichkeitsraum
ω	Realisierung
\emptyset	leere Menge
\int	Integral
$\lim_{x \rightarrow \infty}$	Grenzwert wenn x den Wert unendlich anstrebt
$\rho(x)$	Dichtefunktion

$E(x)$	Erwartungswert von x
$Var(x)$	Varianz von x
σ	Streuung
$F^{-1}(x)$	Umkehrfunktion von $F(x)$

Abkürzungsverzeichnis

AO	Abgabenordnung
BFuP	Betriebswirtschaftliche Forschung und Praxis
BGH	Bundesgerichtshof
Big 4	Die heutigen vier größten Wirtschaftsprüfungsgesellschaften
Big 5	Die ehemals weltweit fünf größten Wirtschaftsprüfungsgesellschaften
BIP	Bruttoinlandsprodukt
BMF	Bundesministerium für Finanzen
BRD	Bundesrepublik Deutschland
DAX	Deutscher Aktienindex
DDR	Deutsche Demokratische Republik
DM	Deutsche Mark
EFG	Entscheidungen der Finanzgerichte
EStG	Einkommensteuergesetz
EY	Ernst & Young Wirtschaftsprüfungsgesellschaft GmbH
EU	Europäische Union
Eurostat	Europäische Kommission für Statistik
FAZ	Frankfurter Allgemeine Zeitung
FG	Finanzgericht
IDEA	Interactive Data Extraction and Analysis
IT	Informationstechnologie
Kfz	Kraftfahrzeug
KS	Kolmogorow-Smirnow (Testverfahren)
NS	Nationalsozialismus bzw. nationalsozialistisch
NSDAP	Nationalsozialistische Deutsche Arbeiterpartei
OECD	Organisation für wirtschaftliche Zusammenarbeit und Entwicklung
PKW	Personenkraftwagen
PLZ	Postleitzahl
RM	Reichsmark
SIT	Sichere Informationstechnologie
SOEP	Sozial Economic Panel
UK	United Kingdom

US	United States
USA	United States of America
VDI	Verband Deutscher Ingenieure
VGR	Volkswirtschaftliche Gesamtrechnung
VKB	Versicherungskammer Bayern
WiSt	Wirtschaft und Statistik
WR	Weimarer Republik
ZRFC	Fachzeitschrift Risk, Fraud & Compliance
Aufl.	Auflage
Abs.	Absatz
Bd.	Band
Bst.	Buchstabe
bzgl.	bezüglich
bzw.	beziehungsweise
ca.	circa
et al.	et alii (und andere)
etc.	et cetera
f.	folgende
ff.	fortfolgende
Fn.	Fußnote
gem.	gemäß
ggf.	gegebenenfalls
ggü.	gegenüber
Hrsg.	Herausgeber
i.A.	im Allgemeinen
i.d.R.	in der Regel
inkl.	inklusive
i.V.m.	in Verbindung mit
Jg.	Jahrgang
Mio.	Millionen
Mrd.	Milliarden
No.	number
Nr.	Nummer

o.g.	oben genannt
s.	siehe
S.	Seite
sog.	sogenannten
u.	und
u.a.	unter anderem
u.U.	unter Umständen
v.a.	vor allem
vgl.	vergleiche
Vol.	Volume
z.B.	Zum Beispiel
z.v.E.	zu versteuerndes Einkommen
Z.	Ziffer
Zif.	Ziffer
zzgl.	zuzüglich

Vorwort

Die vorliegende Arbeit verfasste ich als externer Doktorrand am Fachgebiet für Steuerlehre/Prüfungswesen der Technischen Universität Ilmenau. Sie wurde im August 2016 von der Promotionskommission der Fakultät für Wirtschaft und Medien als Dissertation angenommen und im November 2016 erfolgreich verteidigt.

In diesem Kontext möchte ich der Friedrich-Naumann-Stiftung für die Freiheit besonderen Dank aussprechen, welche mich in dem Zeitraum zwischen Oktober 2014 und November 2016 finanziell und ideell im Rahmen der Promotionsförderung förderte. Ohne diese Unterstützung wäre die Erarbeitung der Dissertationsschrift in dieser Form nicht möglich gewesen.

Gleichzeitig möchte ich mich insbesondere bei meinem Doktorvater, Herrn Univ.-Prof. Dr. Gernot Brähler, für seine ausgezeichnete Unterstützung und sein persönliches Engagement im Rahmen der Betreuung bedanken. Seine immerwährende Diskussionsbereitschaft und seine stets konstruktive Kritik hatten erheblichen Anteil an dem Gelingen der Arbeit. Weiteren Dank möchte ich Herrn Univ.-Prof. Dr. Michael Grüning für die Erstellung des Zweitgutachtens sowie Herrn Univ.-Prof. Dr. Udo Bankhofer, Herrn Dr. Alexander Fox und Herrn Dr. Magnus Richter für ihr Mitwirken in der Promotionskommission aussprechen.

Ilmenau, 30. November 2016

Daniel Brinkert

1. Benford's Law – Eine numerische Untersuchung zur Anwendbarkeit statistischer Anpassungstests und praktische Anwendung auf volkswirtschaftliche Zahlen Deutschlands sowie Kfz-Versicherungsschäden

1.1 Einleitung

„Betrügen und betrogen werden – nichts ist gewöhnlicher auf Erden“, so schrieb bereits der berühmte deutsche Schriftsteller und Dichter Johann Gottfried Seume. Für ihn war eines Gewiss: Der Betrug und die Menschheit sind symbiotisch miteinander verbunden. Ihre Koexistenz war, ist und wird stets gegenwärtig bleiben. So kannten bereits die antiken Griechen in ihrer Mythologie das Geisteswesen *Apate*, welches den Betrug symbolisierte. Bei den alten Römern existierte mit *Fraus* als Pendant ein Gott der Falschheit.¹

Auch in unserer heutigen Gesellschaft sind Manipulationen und bewusste Täuschungen allgegenwärtig: In einer Studie kam die Wirtschaftsprüfungsgesellschaft *KPMG* zu dem Ergebnis, dass durch wirtschaftskriminelle Handlungen in Deutschland, die schwerpunktmäßig durch Abrechnungsbetrug, Bilanzmanipulationen und Steuerstraftaten begangen werden, ein jährlicher Schaden von 40 Mrd. € entsteht.² Allein im Jahr 2014 sind in Deutschland ca. 38.300 Selbstanzeigen aufgrund von Steuerhinterziehung mittels Schwarzgeldkonten im Ausland bei den Strafbehörden eingegangen.³ Im Jahr 2015 waren es immerhin noch rund 20.000 Selbstanzeigen.⁴ Doch nicht nur Privatpersonen und Unternehmer neigen dazu, bewusste Täuschungen durchzuführen. Auch bei Staaten war dies in der Vergangenheit zu beobachten. So gelang es Griechenland mittels der Fälschung eigener makroökonomischer Daten sogar, die Aufnahme in die Euro-Währungsunion zu erwirken. Der hierdurch entstandene Schaden hat mittlerweile kaum abschätzbare Dimensionen erreicht.⁵ Dies sind nur einige Beispiele, deren Aufzählung nahezu beliebig fortgesetzt werden könnte.

¹ Vgl. Von Sybel, Ludwig/ Roscher, Wilhelm Heinrich (Hrsg.), Ausführliches Lexikon der griechischen und römischen Mythologie, Band 1,1, Leipzig 1886, S. 886 f.

² Vgl. KPMG, Wirtschaftskriminalität in Deutschland 2014, <http://www.kpmg.com/de/de/bibliothek/presse/seiten/kpmg-studie-wirtschaftskriminalitaet-in-deutschland-2014.aspx> (abgerufen am 21.01.2015), 2014.

³ Vgl. haufe.de, Rekordjahr für Selbstanzeigen, http://www.haufe.de/steuern/kanzlei-co/steuerhinterziehung-rekordjahr-fuer-selbstanzeigen_170_287304.html (Abgerufen am 13.04.2015), 2015.

⁴ Vgl. Focus Money Online, Weniger Selbstanzeigen von Steuerbetrügern - Druck soll bleiben (Abgerufen am 09.02.2016), 2015.

⁵ Vgl. Manager Magazin, Beitrittsbetrug – Griechenland kommt ohne Strafe davon, <http://www.manager-magazin.de/unternehmen/artikel/a-328064.html> (abgerufen am 21.01.2015), 2004.

So unterschiedlich die Themenfelder der wirtschaftlichen Betrugsfälle sind, so haben sie doch alle eine Gemeinsamkeit: Sie entstehen durch bewusst getätigte falsche Angaben bzw. der Manipulation von Zahlen. Eine effektive Methode, Anhaltspunkte zur Erkennung solcher Manipulationen zu erhalten, stellt die digitale Ziffernanalyse dar.⁶ Diese fasst statistische Testverfahren zusammen, welche die Verteilung einzelner Ziffern eines Datensatzes mit angenommenen hypothetischen Verteilungen vergleichen.⁷ Hierbei unterscheiden sich insbesondere die Verteilungen von manipulierten und nicht manipulierten Zahlen.⁸ Zugrunde gelegt wird hierbei i.A. die sog. „Benford-Verteilung“. Man spricht dann auch von der Anwendung des „Benford-Testes“.

Die logarithmische Benford-Verteilung wurde erstmals durch *Simon Newcomb* im Jahr 1881 entdeckt⁹ und von *Frank Benford* 1938 grundlegend erforscht. Hierbei stellte er fest, dass die Anfangsziffern natürlich vorkommender Zahlen nicht etwa gleichverteilt sind, sondern unterschiedliche Häufigkeiten aufweisen: So kommt die Ziffer eins zu 30,1%, die Ziffer neun hingegen zu lediglich 4,6% vor.¹⁰ Weichen die empirischen Verteilungen der zu analysierenden Datensätze von dieser Verteilung ab, so kann dies ein Indikator dafür sein, dass sich manipulierte Zahlen in diesen befinden.

1.2 Problemstellung

In der Literatur fand der Benford-Test bereits zahlreiche Anwendungen: So wurden bereits makroökonomische Daten, Wahlergebnisse, Rechenschaftsberichte und wissenschaftliche Daten untersucht und die Eignung des Benford-Testes zur Entdeckung fraudulenter Zahlen bestätigt. Auch kann der Benford-Test bei großen Datenmengen leicht angewendet werden, sodass dessen Einsatz im Vergleich zu anderen Prüfungshandlungen sehr kosteneffizient erscheint.¹¹

⁶ Vgl. Watrin, Cristoph/Struffert, Ralf, Benford's Law und Chi-Quadrat-Test – Chancen und Risiken des Einsatzes bei steuerlichen Prüfungen – Zugleich Anmerkungen zum Urteil des FG Münster vom 07.12.2005 1 K 6384/03 E, in: Der Betrieb, 2006, S. 1748.

⁷ Vgl. Odenthal, Roger, Defraudanten auf der Spur, in: Betrieb und Wirtschaft, 2005, S. 1023.

⁸ Vgl. Blenkers, Michael, Chi-Test – oder „Jeder Mensch hat seine Lieblingszahl“, in: Die steuerliche Betriebsprüfung, 2003, S. 261ff.

⁹ Vgl. Newcomb, Simon, Note on the Frequency of Use of the Different Digits in Natural Numbers, in: American Journal of Mathematics, 1881, S. 39, 40.

¹⁰ Vgl. Benford, Frank, The Law of Anomalous Numbers, in: Proceedings of the American Philosophical Society, 1938, S. 554.

¹¹ Vgl. Rauch, Bernhard/ Göttische, Max/ Langenegger, Stephan, Detecting Problems in Military Expenditure Data Using Digital Analysis, in: Defence and Peace Economics, 2014, Vol. 25, No. 2, S. 109.

Jedoch gibt es auch zahlreiche kritische Stimmen, die dessen Anwendbarkeit in Frage stellen.¹² Somit ist die generelle Eignung und Zuverlässigkeit zur Manipulationsaufdeckung bisher nicht eindeutig bewiesen.

Eine sinnvolle und plausible Vorgehensweise, um dieser Fragestellung nachzugehen, wurde hierbei von *Rauch et al.* entwickelt: Sie untersuchten einige Datensätze, von welchen sie sicher wussten bzw. mit einer sehr großen Wahrscheinlichkeit davon ausgehen konnten, dass in ihnen gefälschte Zahlen enthalten sind. So waren beispielsweise die makroökonomischen Daten der EU-Staaten Gegenstand einer Untersuchung. Bei diesen war bereits im Vorfeld bekannt, dass Griechenland Manipulationen an den eigenen Wirtschaftsdaten durchführte, um den Beitritt zu der Euro-Währungsunion zu erwirken. Ausgerechnet diese Daten wiesen jedoch auch die größte Abweichung von der Benford-Verteilung auf.¹³ Daten der Sozialstatistik waren hingegen vollkommen unauffällig.¹⁴ Ein weiteres Untersuchungsprojekt widmete sich der Analyse von Rechenschaftsberichten deutscher Parteien. Auch hier war im Vorfeld der Untersuchung bekannt, dass die Partei „Die Grauen Panther“ Daten fälschte, um die zugewiesenen staatlichen Fördermittel zu erhöhen. Der Benford-Test war hier ebenfalls erfolgreich und identifizierte die größte Deviation bei eben dieser Partei.¹⁵ Auf diese Weise kann der Benford-Test selbst auf seine Tauglichkeit getestet werden: Schlägt er auf die verdächtigen Daten nicht an, so ist er nicht auf die Problemstellung anwendbar. Weisen diese „schwarzen Schafe“ jedoch Abweichungen von der Benford-Verteilung auf, so kann dies als Hinweis für die Wirksamkeit des Benford-Testes verstanden werden. In diesem Kontext ist weiterhin von besonderem Interesse, ob weitere Untersuchungseinheiten Auffälligkeiten zeigen.

Dieser Vorgehensweise folgend erscheint die Anwendung des Benford-Testes auch auf historische makroökonomische Daten Deutschlands als sinnvoll. Während die Wirtschaftsdaten der Bundesrepublik Deutschland wenig Plausibilitäten für manipuliertes Datenmaterial vorweisen, existierten mit der Diktatur des Nazi-Regimes (im Folgenden: „NS-Regime“) in der Zeit von

¹² Eine detaillierte Darlegung der bisherigen Forschungsprojekte erfolgt in Kapitel 2.

¹³ Vgl. Rauch, Bernhard/ Götttsche, Max/ Brähler, Gernot/ Engel, Stefan, Fact and Fiction in EU-Governmental Economic Data, in: German Economic Review, 2011, S. 243-255.

¹⁴ Vgl. Rauch, Bernhard/ Götttsche, Max/ Brähler, Gernot/ Kronfeld, Thomas, Deficit versus social statistics: empirical evidence for the effectiveness of Benford's law, in: Applied Economic Letters, 2014, Vol. 21, No. 3, S. 147-151.

¹⁵ Vgl. Rauch, Bernhard/ Götttsche, Max/ Brähler, Gernot/ Geidel, Felix/ Pietras, Thomas, Überprüfung der Rechenschaftsberichte deutscher Parteien mit Hilfe des Benford's Law, in: BFuP, 2014, S. 175-191.

1933 bis 1945 und der Deutschen Demokratischen Republik (1949 – 1990) gleich zwei Staatsformen, in denen Falschangaben in den veröffentlichten Wirtschaftsdaten als sehr wahrscheinlich und durchaus realistisch angenommen werden können.

Während des Dritten Reiches begann unmittelbar nach der Machtergreifung durch Adolf Hitler und seine NSDAP ein strategischer Umschwung in der Wirtschaftspolitik, der die systematische Aufrüstung der Wehrmacht als Vorbereitung auf den geplanten zweiten Weltkrieg beinhaltete.¹⁶ Die Regierung hatte jedoch allen Grund dazu, die ökonomischen Auswirkungen dieser äußerst kostspieligen Aufrüstung vor den anderen Regierungen und dem eigenen Volk so lange wie möglich geheim zu halten, um sich selbst sowie die Regierungsentscheidungen propagandistisch so positiv wie möglich darzustellen und so das Volk hinter sich zu wissen. Auch die ab 1938 in großem Stil durchgeführte Zwangsarbeit der Juden, Sinti und Roma und deren Einsatz in der Rüstungsindustrie hätten sicherlich ausreichend Motivation begründet, verheimlicht zu werden.¹⁷

In der Deutschen Demokratischen Republik pries sich die Regierung nach außen v.a. mit der vorhandenen Vollbeschäftigung, die für jeden bezahlbaren Mieten sowie einer guten medizinischen Versorgung. Die Kehrseite der Medaille zeigte sich jedoch in einer prekären Arbeitsproduktivität, einer mangelhaften Investitionsquote sowie einem hohen Verschuldungsgrad durch fehlende Exporte. Diese Umstände machten es langfristig unmöglich, das vorhandene Sozialsystem finanziell zu tragen.¹⁸ In den zum damaligen Zeitpunkt geheimen sog. „Schürer-Papieren“ des Leiters der Zentralen Plankommission wies Gerhard Schürer gar auf einen bevorstehenden Staatsbankrott der DDR hin.¹⁹ Möglicherweise versuchte die DDR-Führung also während einer langen Zeitspanne, die eigene finanzielle Situation nach außen zu beschönigen. Auch wurde durch die Regierung mit einer hohen Zustimmungsrate der Bevölkerung geworben, auf welche durch die eindeutig erscheinenden Wahlergebnisse geschlossen wurde. Jedoch ist längst erwiesen, dass es hierbei zu systematischen Wahlmanipulationen gekommen ist. Nach heutigen

¹⁶ Vgl. Bundeszentrale für politische Bildung, Wirtschaft und Gesellschaft unterm Hakenkreuz, <http://www.bpb.de/geschichte/nationalsozialismus/dossier-nationalsozialismus/39551/wirtschaft-und-gesellschaft?p=all> (Abgerufen am 28.03.2015), 2005.

¹⁷ Vgl. Süddeutsche Zeitung, Zwangsarbeit im Dritten Reich, <http://www.sueddeutsche.de/politik/deutschland-ein-sklavenhalterstaat-zwangsarbeit-im-dritten-reich-1.526445> (Abgerufen am 28.03.2015), 2008.

¹⁸ Vgl. Die Bundesregierung, Die wirtschaftliche Lage der DDR vor der Wende 1989, <http://www.bundesregierung.de/Content/DE/Magazine/MagazinWirtschaftFinanzen/072/s1-die-wirtschaftliche-lage-der-ddr-vor-der-wende.html> (Abgerufen am 28.03.2015), 2009.

¹⁹ Vgl. Mitteldeutscher Rundfunk, Wie pleite war die DDR?, <http://www.mdr.de/damals/ddrpleite104.html> (Abgerufen am 28.03.2015), 2011.

Erkenntnissen wurden vorsätzliche Falschauszählungen der Stimmen sowie nachträgliche Korrekturen der Ergebnisse durchgeführt, die die Gesamtergebnisse um ca. neun bis neunzehn Prozentpunkte verfälschten.²⁰ Es liegt also nahe, dass mit den Wirtschaftsdaten ähnlich verfahren worden ist.

Hieraus ergibt sich die grundlegende Vermutung, dass die Wirtschaftsdaten des Dritten Reiches und der DDR im Vergleich zu denjenigen der Bundesrepublik Deutschland eine deutlich höhere Abweichung von der Benford-Verteilung aufweisen. Dies gilt auch deshalb, da es sich sowohl bei dem NS-Regime als auch bei der DDR-Führung um rücksichtslose Regierungen handelte, denen es neben den bestehenden Anreizen an jeglicher Moral zur Ehrlichkeit fehlt. Denn wer keinen Respekt vor dem menschlichen Leben zeigt und ganze Ethnien systematisch ermordet²¹ oder die eigene Bevölkerung an der Ausreise mit tödlichem Schießbefehl hindert²², wird auch keine Hemmungen vor der Manipulation eigener Wirtschaftsdaten haben. Dies müsste bei entsprechender Eignung durch die Ergebnisse des Benford-Testes eindeutig belegt werden.

Doch nicht nur in historischen Wirtschaftsdaten gelten Manipulationen als wahrscheinlich. In der Versicherungsbranche entgehen den Versicherungen erwiesenermaßen jährlich mehr als vier Milliarden Euro in Folge des Betrugs durch Versicherungsnehmer – und dies allein in dem Bereich der Unfall- und Schadenversicherung.²³ Ein besonders hoher Anteil von 1,2 Milliarden Euro entfällt hierbei auf Kfz-Schäden. Bei diesen wird davon ausgegangen, dass in mehr als 10% der Fälle Manipulationen vorgenommen werden, um den an die Versicherungsnehmer auszuzahlenden Betrag zu erhöhen. Der Tatbestand geht sogar so weit, dass Unfälle absichtlich herbeigeführt sowie mit Personen aus dem eigenen Bekanntenkreis abgesprochen werden.²⁴ Für die Versicherungen hat dies zur Konsequenz, dass nahezu jeder Einzelschaden von Sachverständigen intensiv geprüft und Positionen in den eingereichten Rechnungen herausgestrichen

²⁰ Vgl. Mitteldeutscher Rundfunk, Blockparteien und Einheitslisten – Wahlen in der DDR, http://www.mdr.de/damals/eure-geschichte/themen/wahlen106_page-0_zc-6615e895.html (abgerufen am 28.03.2015), 2012.

²¹ Vgl. Bundeszentrale für politische Bildung, Massenmord und Holocaust, <http://www.bpb.de/izpb/151942/massenmord-und-holocaust> (Abgerufen am 09.02.2016), 2012.

²² Vgl. Bundeszentrale für politische Bildung, Wer trägt die Schuld? – Schießbefehl und Mauertote, <http://www.bpb.de/geschichte/deutsche-geschichte/kontraste/42470/schiessbefehl-und-mauertote> (Abgerufen am 09.02.2016), 2012.

²³ Vgl. Gesamtverband der Deutschen Versicherungswirtschaft, Scherpunkt Versicherungsbetrug, <http://www.gdv.de/versicherungsbetrug> (abgerufen am 21.01.2015), 2014.

²⁴ Vgl. Gesamtverband der Deutschen Versicherungswirtschaft, Versicherungsbetrug in der Kfz-Versicherung: Nicht jeder Unfall ist Zufall – Provozierte Verkehrsunfälle, <http://www.gdv.de/2008/04/versicherungsbetrug-in-der-kfz-versicherung-nicht-jeder-unfall-ist-zufall-provozierte-verkehrsunfaelle> (Abgerufen am 28.03.2015), 2008.

werden, was einen enormen Aufwand für die Versicherungskonzerne zur Folge hat.²⁵ Wendet man den Benford-Test auf Schadensfälle der Versicherungen an, ist zu prüfen, ob dieser demnach ebenfalls dazu in der Lage ist, die durchgeführten Falschangaben aufzudecken.

Neben der Fragestellung, ob sich der Benford-Test zur Manipulationsaufdeckung eignet, ist nach derzeitigem Forschungsstand jedoch auch ungeklärt, wie eine mögliche Abweichung von der Benford-Verteilung am besten und effizientesten aufgedeckt werden kann, sprich welche statistischen Tests sich in welchem Maße zur Anwendung innerhalb des Benford-Tests eignen. Hierbei unterscheiden sich die Meinungen in der Literatur zum Teil deutlich. So greift ein Großteil der Forschungsgruppen wie z.B. *Bauer und Gross*²⁶, *Cleary und Thibodeau*²⁷, *Diekmann*²⁸, *Günnel und Tödter*²⁹ sowie *Gonzalez-Garcia und Pastor*³⁰ im Rahmen der Ziffernanalyse beispielsweise auf den sog. „Chi-Quadrat-Anpassungstest“ zurück. Dieser Signifikanztest ist zwar sehr einfach zu handhaben, birgt jedoch gleichzeitig die Gefahr von Fehlinterpretationen. Dies gilt v.a. für sehr große Stichprobenumfänge, da in diesem Fall bereits bei kleinsten Abweichungen eine signifikante Deviation attestiert wird und somit falsch positive Ergebnisse hervorgebracht werden.³¹ Demnach ist auch ein Ergebnisvergleich zwischen Datensätzen unterschiedlicher Stichprobengrößen nur eingeschränkt möglich. *Rauch et al.* umgehen diese Problematik, indem sie die Testgröße des Chi²-Anpassungstestes lediglich neben weiteren Abstandsmaßen als Bewertungskriterium heranziehen und auf Basis dessen eine Reihenfolge der Untersuchungsobjekte hinsichtlich der Benford-Abweichung bilden. Andere Forscher wie z.B.

²⁵ Vgl. Focus Online, Streichen mit System, http://www.focus.de/finanzen/versicherungen/kfz-versicherung/recht-streichen-mit-system_aid_220818.html (Abgerufen am 29.03.2015), 2007.

²⁶ Vgl. Bauer, Johannes/ Gross, Jochen, Difficulties in Detecting Fraud? The Use of Benford's Law on Regression Tables, in: *Journal of Economics and Statistics*, 2011, S. 744-748.

²⁷ Vgl. Cleary, Richard/ Thibodeau, Jay C., Applying Digital Analysis Using Benford's Law to Detect Fraud: The Dangers of Type I Errors, in: *Auditing: A Journal Of Practice & Theory*, Vol. 24, No. 1, 2005, S. 79.

²⁸ Vgl. Diekmann, Andreas, Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data, in: *Journal of Applied Statistics*, 2007, S. 324.

²⁹ Vgl. Günnel, Stefan/ Tödter, Karl-Heinz, Does Benford's Law hold in economic research and forecasting?, in: *Empirica*, 2009, S. 285.

³⁰ Vgl. Gonzalez-Garcia, Jesus/ Pastor, Gonzalo, Benford's Law and Macroeconomic Data Quality, in: *International Monetary Fund, Working Paper*, 2009, S. 7-10.

³¹ Vgl. Rodriguez, Ricardo J., Reducing False Alarms in the Detection of Human Influence on Data, in: *Journal of Accounting, Auditing & Finance*, 2004, S. 141-158.

*Carslaw*³² sowie *Busta/Weinberg*³³ und *Möller*³⁴ greifen zur Beurteilung der Datenqualität in erster Linie auf die sog. „Z-Statistik“ zurück. Die Wissenschaftler *Cho/Gaines*³⁵, *Günnel/Tödter*³⁶ und *Giles*³⁷ verwenden den KS- bzw. Kuiper-Test. *Nigrini* entwickelte gar ein eigenes Verfahren, das sog. „Verzerrungsfaktor-Modell“.³⁸

Folglich existiert derzeit aufgrund der Vielzahl an statistischen Tests und Maßzahlen keine einheitliche Vorgehensweise bei der Bestimmung der Benford-Abweichung. Somit ist zum einen die Vergleichbarkeit der Ergebnisse verschiedener Untersuchungsprojekte nicht gewährleistet. Zum anderen wirft dies die Frage auf, welche der Vorgehensweisen sich am besten zur Anwendung im Rahmen des Benford-Testes eignet bzw. welche Verfahren sich unter bestimmten Bedingungen am besten bewähren.

1.3 Zielstellung

Ziel dieser Arbeit ist es, die Effektivität und die Einschlägigkeit des Benford-Testes zu überprüfen. Zu diesem Zweck erfolgt dessen Anwendung auf historische Wirtschaftsdaten Deutschlands der Jahre 1918 bis 1990. Hierbei ist von besonderem Interesse, ob durch die Ergebnisse eine Unterschiedlichkeit der Datenqualität im Falle der verschiedenen Staatsformen, die während dieser Zeitspanne existierten, angezeigt wird. Da vor allem die Regierungen des Nationalsozialismus in der Zeit von 1933 und 1945 sowie das Regime der Deutschen Demokratischen Republik (1949 – 1990) größte Anreize zur Datenmanipulation gehabt hätten sowie durch ihre verbrecherischen Taten ihre fehlende Ehrlichkeit und nicht vorhandene Hemmschwelle gegenüber unmoralischen Handlungen unter Beweis stellten, sollte der Benford-Test die grundsätzliche Eignung besitzen, Manipulationen in deren Datensätzen aufzudecken. Im Vergleich dazu

³² Vgl. Carslaw, Charles A. P. N., Anomalies in Income Numbers: Evidence of Goal Oriented Behaviour, in: The Accounting Review, 1988, S. 323.

³³ Vgl. Busta, Bruce/ Weinberg, Randy, Using Benford's law and neural networks as review procedure, in: Managerial Auditing Journal, 1998, S. 362.

³⁴ Vgl. Möller, Manuela, Measuring the quality of auditing services with the help of Benford's Law?, <http://ssrn.com/abstract=1529307> (Abruf vom 23.01.2015), 2009.

³⁵ Vgl. Cho, Wendy K. Tam/ Gaines, Brian J., Breaking the (Benford) Law: Statistical Fraud Detection in Campaign Finance, in: The American Statistician, 2007, S. 221-223.

³⁶ Vgl. Günnel, Stefan/ Tödter, Karl-Heinz, Does Benford's Law hold in economic research and forecasting?, in: Empirica, 2009, S. 279.

³⁷ Vgl. Giles, David E., Benford's Law and naturally occurring prices in certain eBay auctions, in: Applied Economic Letters, 2007, S. 160.

³⁸ Vgl. Nigrini, Mark, A Taxpayer Compliance Application of Benford's Law, in: JATA, 1996, Vol. 18, Nr. 1, S. 76.

sollte dieser bei Zahlen der Bundesrepublik Deutschland auf eine deutlich höhere Datenqualität schließen lassen. Dieser These wird im Rahmen der vorliegenden Arbeit nachgegangen.

Die zweite empirische Anwendung des Benford-Testes erfolgt auf Kfz-Schadensfälle eines Versicherungsunternehmens. Auch hier wird geprüft, ob der Benford-Test dazu in der Lage ist, die fraudulent Zahlen aufzuspüren. Es wird also die generelle Anwendbarkeit des Benford-Testes auf Versicherungsdaten auf die Probe gestellt und die Grenzen des Benford-Testes aufgezeigt. Im Zuge dessen wird der Datensatz hinsichtlich spezieller Kriterien, die in Abschnitt 5.2 näher erläutert werden, weiter aufgegliedert und verifiziert, ob die jeweils höchsten Benford-Deviations plausibel begründet werden können. Bei erfolgreicher Anwendung des Benford-Testes könnten Versicherungsunternehmen das Testverfahren insbesondere dazu verwenden, Anhaltspunkte dafür zubekommen, welche Schadensfälle mit einer besonders großen Manipulationswahrscheinlichkeit einhergehen und in diesem Zusammenhang einer detaillierteren Prüfung unterzogen werden sollten.

In beiden Untersuchungen wird zunächst geprüft, ob der Datensatz die grundlegenden Anforderungen an eine Benford-Verteilung erfüllt und somit einer Anwendbarkeit des Benford-Testes nichts entgegensteht. Mittels der als geeignet eingestuften Testverfahren erfolgt dann die Ermittlung der Deviation von der Benford-Verteilung. Im Anschluss werden die Ergebnisse auf ihre Plausibilität hin untersucht und kritisch ausgewertet. Auffällige Abweichungen können im Zuge dessen als Indiz für vorhandene Manipulationen interpretiert werden. Jedoch lassen diese keinesfalls verbindliche Aussagen in Form von Beweisen zu. Aus diesem Grund ist es generell ratsam, eine effiziente Prüfreihefolge anhand der Benford-Deviation zu bilden, welche Ausgangspunkt für eine detailliertere Prüfung der Daten darstellt.

Als Unterziel wird im Vorfeld der Fragestellung nachgegangen, auf welche Weise eine auftretende Benford-Deviation am effizientesten aufgedeckt werden kann. Hierzu erfolgt eine vergleichende Analyse der statistischen Tests und Maßzahlen, auf dessen Basis eine Empfehlung zur Anwendung eines geeigneten Instrumentariums bzw. geeigneter Instrumentarien ausgesprochen und Aussagen getroffen werden, welche der Verfahren sich im Falle bestimmter Stichprobeneigenschaften besonders zur Manipulationsaufdeckung eignen. Somit können zukünftig Testverfahren und Maßzahlen zielgerichteter verwendet werden. Untersuchungsergebnisse lassen dann konkretere Schlüsse zu und ermöglichen es somit, Manipulationen in Datensätzen effizienter aufzudecken als dies bislang möglich ist. Dies ist eine wesentliche Vorausset-

zung, um die Praxistauglichkeit der Ziffernanalyse unter Verwendung des Benfordschen Gesetzes signifikant zu erhöhen. Dies gilt v.a. dann, wenn der Datensatz Inhomogenitäten bzgl. der Eigenschaften wie z.B. dem Stichprobenumfang oder dem Grad der Abweichungen aufweist. Des Weiteren ist es durch die einheitliche Nutzung geeigneter Verfahren möglich, auch Ergebnisse unterschiedlicher Untersuchungen zu vergleichen. Die Erkenntnisse aus diesem Unterziel werden anschließend auf die zwei Anwendungsbeispiele angewandt.

1.4 Gang der Arbeit

Die Arbeit wird insgesamt in zwei wesentliche Teile untergliedert: Während sich Kapitel zwei, drei und vier zunächst theoretisch mit der digitalen Ziffernanalyse auseinandersetzen, erfolgen in Kapitel fünf und sechs die empirischen Anwendungen des Benford-Testes.

Abschnitt zwei legt zunächst den bisherigen Forschungsstand zu der Thematik des Benford's Law dar. Hierbei wird der inhaltliche Schwerpunkt zunächst auf die Bereiche der Rechnungslegung, Steuererklärung und –prüfung gelegt. Im Anschluss werden Forschungsarbeiten in den Thematiken der makroökonomischen und statistischen Daten zusammengefasst. Abschließend werden weitere Anwendungsgebiete wie z.B. Aktienkurse oder der Betrugsaufdeckung bei Wahlen aufgezeigt.

Das dritte Kapitel steht ganz im Fokus der theoretischen Grundlagen des Benfordschen Gesetzes sowie den im Rahmen dieser Forschungsarbeit näher analysierten statistischen Testverfahren und Maßzahlen. So wird in Abschnitt 3.1 zunächst ein Einblick in die historische Entwicklung des Benfordschen Gesetzes gegeben und diese allgemein erläutert. Die Voraussetzungen, die an einen Benford-verteilten Datensatz gestellt werden, sowie die Eigenschaften, die ein solcher mit sich bringt, werden in Abschnitt 3.2 beschrieben. Abschnitt 3.3 widmet sich kritisch der Fragestellung, ob und unter welchen Annahmen sich das Benfordsche Gesetz zur Manipulationsaufdeckung eignet. Im Anschluss werden sämtliche eingesetzte und betrachtete statistische Testverfahren und Kennzahlen vorgestellt und deren theoretische Grundlagen ausführlich beschrieben.

Kapitel 4 dieser Arbeit widmet sich der numerischen Analyse der Arbeit, mittels welcher die Eigenschaften der unterschiedlichen Testverfahren und Maßzahlen untersucht und verglichen werden. Hierzu wird in Abschnitt 4.1 zunächst der Grundgedanke der Untersuchung näher er-

läutert sowie die Methodik und Vorgehensweise beschrieben. Da zur Generierung der im Rahmen der Analyse verwendeten Datensätze die Monte-Carlo-Simulation eingesetzt wurde, wird das Simulationsprinzip in Abschnitt 4.2 ausführlich erklärt. In Abschnitt 4.3 folgt dann die Darlegung der gewonnenen Ergebnisse für jedes der untersuchten Verfahren. Abschnitt 4.4 stellt auf Basis dieser Ergebnisse die Eigenschaften und Anforderungen der einzelnen Testverfahren und Maßzahlen dar. Gleichzeitig werden Empfehlungen ausgesprochen, unter welchen Umständen sich die entsprechenden Verfahren eignen sowie wann von einer Anwendung besser Abstand genommen werden sollte.

In dem fünften Kapitel dieser Arbeit erfolgt die erste empirische Untersuchung auf historische makroökonomische Daten Deutschlands der Staatsformen der Weimarer Republik, des NS-Regimes, der DDR sowie der Bundesrepublik Deutschland zwischen 1918 und 1990. In diesem Kontext erfolgt in Abschnitt 5.1 zunächst eine grundlegende Beschreibung des Datensatzes. Anschließend wird detailliert beschrieben, wie die Daten untersucht werden und die grundlegenden Anforderungen an den Datensatz überprüft. Abschnitt 5.3 widmet sich darauf aufbauend der ausführlichen Ergebnisdarstellung des aggregierten Datensatzes, der einzelnen Staatsformen sowie einer Einzeljahresauswertung. In dem vierten Abschnitt des fünften Kapitels werden die in 5.3 gewonnenen Ergebnisse auf ihre Plausibilität hin überprüft und bewertet.

Im sechsten Kapitel erfolgt die Applikation des Benford-Testes auf Versicherungsdaten von Kfz-Unfallschäden. Auch hier erfolgt zunächst in Abschnitt 6.1 eine genaue Beschreibung des Datensatzes, ehe in Abschnitt 6.2 das Untersuchungsdesign vorgestellt und mögliche Restriktionen identifiziert werden. Ebenso erfolgt die Überprüfung der allgemeinen Voraussetzungen des Datensatzes an die Benford-Verteilung. In dem dritten Abschnitt des sechsten Kapitels werden die Ergebnisse der Benford-Analyse des aggregierten Gesamtdatensatzes sowie der zwei betrachteten Teildatensätze detailliert dargestellt. Im Anschluss werden die Ergebnisse in Abschnitt 6.4 bewertet und einem Plausibilitätscheck unterzogen.

Im siebten Kapitel werden die im Rahmen dieser Forschungsarbeit gewonnen Erkenntnisse zusammengefasst und ein Ausblick auf potenzielle zukünftige Forschungsfelder gegeben.

2. Aktueller Stand der Forschung zu Benford's Law

2.1 Anwendung in der Rechnungslegung sowie bei Steuererklärungen und –prüfungen

Bereits im Jahr 1997 legten *Nigrini* und *Mittermaier* die wesentlichen Instrumente zur Applikation der Ziffernanalyse dar und wendeten diese im Rahmen einer empirischen Studie an. Sollten Fälschungen auftreten, erwarteten sie in erster Linie Duplikate und Rundungen, weshalb sie bei einer genauen Analyse die Untersuchung der ersten Ziffern, der zweiten Ziffern und der ersten beiden Ziffern in Kombination empfehlen.³⁹

Im Rahmen der Applikation des Benford-Testes in der Rechnungslegung von Unternehmen, kam dieser in erster Linie zum Einsatz, um mögliche systematische Einflussnahmen auf Gewinnzahlen eines Unternehmens, die auch als „Earnings Management“ bezeichnet werden, aufzudecken.⁴⁰ Zum einen könnte hierbei der Druck des Erreichens interner Ziele ursächlich für die Motivation zur systematischen Datenanpassung sein.⁴¹ Doch auch die externe Wirkung der Bilanzdaten impliziert eine besondere Eignung der Ziffernanalyse aufgrund der sog. „cognitive reference points“. Diese beschreiben die Tatsache, dass sich Menschen bei der Wahrnehmung einer Zahl deren erste Ziffer inklusive der Zehnerpotenz merken. Alle weiteren Stellen sind lediglich von nebensächlicher Bedeutung. Dies hat zur Folge, dass z.B. die Zahl 1.369,14 als 1.000 und die 21.897 als 20.000 im Gedächtnis bleibt.⁴² Zeitgleich hat dies auch zur Konsequenz, dass Manipulationen bzw. systematische Aufrundungen in erster Linie an den ersten Stellen durchgeführt werden, da der erzielte Effekt hierbei am größten ist.

Im Jahr 1988 gelang es dem Wissenschaftler *Carslaw*, mittels der Ziffern-Analyse einen ersten Beweis für systematische Aufrundungen im Falle von Gewinnzahlen neuseeländischer Unternehmen zu liefern: Er stellte bei der Analyse der zweiten Ziffern fest, dass zu viele Nullen und zu wenige Neunen vorhanden sind, was er als eindeutiges Muster des Aufrundens interpretiert.⁴³ Der Forscher *Thomas* konnte zudem zeigen, dass im Falle amerikanischer Unternehmen

³⁹ Vgl. Nigrini, Mark J./ Mittermaier, Linda J., *The Use of Benford's Law as an Aid in Analytical Procedures*, 1997, S. 58-64.

⁴⁰ Vgl. Healy, Paul M./ Wahlen, James M., *A Review of the Earnings Management Literature and Its Applications for Standard Setting*, in: *Accounting Horizons*, 1999, S. 368.

⁴¹ Vgl. Skousen, Christopher J./ Guan, Liming/ Wetzels, T. Sterling, *Anomalies and Unusual Patterns in Reported Earnings: Japanese Managers Round Earnings*, in: *Journal of International Financial Management and Accounting*, 2004, S. 232.

⁴² Vgl. Niskanen, Jyrki/ Keloharju, Matti, *Earnings cosmetics in a tax-driven accounting environment: evidence from Finnish public firms*, in: *The European Accounting Review*, 2000, S. 444.

⁴³ Vgl. Carslaw, Charles A. P. N., *Anomalies in Income Numbers: Evidence of Goal Oriented Behaviour*, in: *The Accounting Review*, 1988, S. 322-324.

dies nur bei positiven Gewinnzahlen zu beobachten ist. Bei Verlusten treten hingegen keinerlei Rundungsmuster auf.⁴⁴ Doch nicht nur im Falle amerikanischer und neuseeländischer Unternehmen konnten die Rundungsmuster bestätigt werden, sondern auch für Firmen in Großbritannien⁴⁵, Finnland⁴⁶ und Japan⁴⁷. In einer weiteren Untersuchung konnte dieses Phänomen durch *Kinnunen* und *Koskela* für mehr als 22.000 Unternehmen in 18 Staaten ebenfalls verifiziert werden.⁴⁸ Wurden in den vorherigen Untersuchungen weitestgehend jährliche Daten analysiert, so verglichen *Guan et al.* die Datenqualität einzelner Quartale. Auffällig war hierbei, dass v.a. die Daten des letzten Quartals eine im Vergleich deutlich bessere Benford-Konformität aufwiesen, was sie mit der unmittelbar anstehenden Examinierung durch Wirtschaftsprüfer begründeten: Aufgrund der gegenwärtigen Gefahr, die Rundungen könnten entdeckt werden, gehen die Mitarbeiter im letzten Quartal eines Jahres bei der Datenangabe weniger risikofreudig vor.⁴⁹

Diese Erkenntnisse bewegten *van Caneghem* dazu, den Einfluss von Wirtschaftsprüfern auf die Datenqualität der Bilanzdaten näher zu untersuchen. Er examinierte Vorsteuerkennzahlen und verglich dabei Daten von Unternehmen, die durch eine der großen damaligen „Big-5“-Prüfungsgesellschaften geprüft wurden mit denjenigen, die von einer kleineren Gesellschaft examiniert wurden. Prinzipiell konnte er zeigen, dass im Falle der Prüfung durch kleinere Kanzleien eine schlechtere Benford-Konformität zu beobachten ist. Bei Betrachtung des Spezialisierungsgrades der Prüfer stellte er zudem fest, dass mit ansteigendem Spezialisierungsgrad auch der Anteil der gerundeten Daten abnimmt.⁵¹ Eine ähnliche Analyse führte *Möller* im Falle deutscher DAX-Konzerne durch. Er bestätigte grundsätzlich, dass die Größe einer Prüfungsgesellschaft positiv mit der Qualität der Bilanzdaten korreliert: Diejenigen Unternehmen, welche von

⁴⁴ Vgl. Thomas, Jacob K., Unusual Patterns in Reported Earnings, in: *The Accounting Review*, 1989, S. 787.

⁴⁵ Vgl. Van Caneghem, Tom, Earnings Management induced by cognitive reference points, in: *British Accounting Review*, 2002, S. 174-176.

⁴⁶ Vgl. Skousen, Christopher J./ Guan, Liming/ Wetzel, T. Sterling, Anomalies and Unusual Patterns in Reported Earnings: Japanese Managers Round Earnings, in: *Journal of International Financial Management and Accounting*, 2004, S. 229.

⁴⁷ Vgl. Niskanen, Jyrki/ Keloharju, Matti, Earnings cosmetics in a tax-driven accounting environment: evidence from Finnish public firms, in: *The European Accounting Review*, 2000 S. 449-451.

⁴⁸ Vgl. Kinnunen, Juha/ Koskela, Markku, Who Is Miss World in Cosmetic Earnings Management? A Cross-National Comparison of Small Upward Rounding of Net Income Numbers among Eighteen Countries, in: *Journal of International Accounting Research*, 2003, S. 39-68.

⁴⁹ Vgl. Guan, Liming/ He, Daoping/ Yang, David, Auditing, integral approach to quarterly reporting, and cosmetic earnings management, in: *Managerial Auditing Journal*, 2006, S. 579.

⁵⁰ Die ehemals fünf größten Wirtschaftsprüfungsgesellschaften.

⁵¹ Vgl. Van Caneghem, Tom, The Impact of Audit Quality on Earnings Rounding-up Behaviour: Some UK Evidence, in: *European Accounting Review*, 2004, S. 775-781.

einer „Big-4“⁵² -Gesellschaft revidiert wurden, wiesen eine bessere Annäherung an die Benford-Verteilung auf.⁵³

Im Jahr 2003 fand der Benford-Test erstmals eine konkrete Forschungs-Anwendung in der Jahresabschlussprüfung. *Quick* und *Wolz* zeigten in diesem Zusammenhang zunächst, dass die ersten beiden Ziffern in Jahresabschlüssen der Benford-Verteilung folgen. Daraus schlussfolgerten sie die grundsätzliche Eignung der Ziffernanalyse zur Aufdeckung von Manipulationen in diesem Anwendungsgebiet. Um tatsächliche Falschangaben zu beweisen, sei es jedoch erforderlich, den Benford-Test um die klassischen Prüfungshandlungen zu erweitern.⁵⁴ Ergänzend zu dieser Forschungsarbeit zeigte *Rodriguez*, dass Nettogewinn, Bilanzsumme und Umsatz eine Benford-Verteilung aufweisen.⁵⁵ Auch *Henselmann et al.* bestätigten die Eignung des Benford-Testes als ein sinnvolles Instrumentarium zur Aufdeckung von Manipulationen in Jahresabschlüssen.⁵⁶

Die indischen Forscher *Kuntluru et al.* untersuchten die Gewinn- und Verlustrechnungen von 22.147 nicht an der Börse gelistete Unternehmen und zeigten, dass deren Datenqualität tendenziell schlechter ist als derjenigen, deren Aktien offiziell an der Börse gehandelt werden.⁵⁷

Watrin und *Ullmann* bezogen sich in einer ihrer Untersuchungen auf Cashflow-Daten deutscher Unternehmen. Bei diesen erwarteten sie grundsätzlich nur geringe Abweichungen, was sie mittels einer empirischen Untersuchung auch bestätigten.⁵⁸

Im Jahr 2014 gelang es *Mollat* im Rahmen einer Kooperation zwischen der Steuerberatungs- und Wirtschaftsprüfungsgesellschaft Ernst & Young und dem Fraunhofer-Institut für Sichere Informationstechnologie, eine Weiterentwicklung der Ziffernanalyse zu veröffentlichen. Bei

⁵² Die heutigen vier größten Wirtschaftsprüfungsgesellschaften.

⁵³ Vgl. Möller, Manuela, Measuring the quality of auditing services with the help of Benford's Law?, <http://ssrn.com/abstract=1529307> (Abruf vom 27.01.2015), 2009.

⁵⁴ Vgl. Quick, Reiner/ Wolz, Matthias, Benford's Law in deutschen Rechnungslegungsdaten, in: Betriebswirtschaftliche Forschung und Praxis, 2003, S. 221-223.

⁵⁵ Vgl. Rodriguez, Ricardo J., First Significant Digit Patterns from Mixtures of Uniform Distributions, in: The American Statistician, 2004, S. 68, 69.

⁵⁶ Vgl. Henselmann, Klaus/ Scherr, Elisabeth/ Ditter, Dominik, Applying Benford's Law to individual financial reports: An empirical investigation on the basis of SEC XBRL filings, Nürnberg 2012, S. 27.

⁵⁷ Vgl. Kuntluru, Sudershan/ Shette, Rachappa/ Achalapathi, K. V., Applying Benford's Law to Examine the Quality of Reported Income Numbers of Unlisted Firms, in: International Journal of Economics and Finance, Vol. 7, No. 12, 2015, S. 1.

⁵⁸ Vgl. Watrin, Christoph/ Ullmann, Robert, Ziffernanalyse in der steuerlichen Betriebsprüfung: Voraussetzungen, Funktionsweise, Anwendungsmöglichkeiten, in: Die Wirtschaftsprüfung, 2009, S. 104, 105.

der herkömmlichen Form kommt es z.B. aufgrund vorhandener Preis-Obergrenzen oder Mindestwerten zu falsch positiven Ergebnissen. Die weiterentwickelte Methode hingegen entwirft auf Basis der vorliegenden Daten ein eigenes Modell, welches die Verteilung simuliert. Hierbei werden insbesondere verwandte Verteilungen aus Mathematik und Statistik mittels komplexer Methoden integriert und so die Anwendungsfelder systematisch erweitert. Aufgrund der großen Effizienz und den geringen Kosten, stellt dieses Modell die Basis dar, um im Alltag forensischer Wirtschaftsprüfer Anwendung zu finden.⁵⁹

Auch im Rahmen von Steuererklärungen fand das Benfordsche Gesetz bereits Anwendungen. So wendete *Nigrini* im Jahr 1996 die Ziffernanalyse zur Untersuchung amerikanischer Steuererklärungen an. Hierbei fiel ihm insbesondere eine hohe Benford-Abweichung bei Geringverdienern mit einem großen Steuer-Rückerstattungspotential auf.⁶⁰ *Posch* untersuchte 2003 Steuererklärungen eines nordrheinwestfälischen Finanzamtes, um die Qualität verschiedener Testverfahren zu vergleichen. Insbesondere führte er eine neue Klasse sog. „Invarianztests“ ein, welche sich v.a. im Falle unterschiedlicher Stichprobenumfänge eignen sollen.⁶¹

Auch Betriebsprüfer fanden in der Vergangenheit bereits Anwendungsfelder der Ziffernanalyse unter Verwendung des Benford's Law: So ist der Chi²-Anpassungstest⁶² in die Software IDEA integriert, welche bei Betriebsprüfern standardmäßig zur Anwendung kommt.⁶³ Deutsche Gerichte ziehen Resultate der Ziffernanalyse ebenfalls mit zur Urteilsfindung heran. Die Zulässigkeit bestätigte der BGH bereits im Jahr 1989 in einem Urteil.⁶⁴ Das FG Münster nutzte die Resultate des Chi²-Anpassungstestes in zwei Fällen⁶⁵, machte jedoch deutlich, dass diese lediglich als Indiz dienen und keine Beweisfunktion haben.⁶⁶

⁵⁹ Vgl. Mollat, Ingolf, Benford reloaded – IT-Forensik: EY und Fraunhofer SIT gelingt der Durchbruch, in: ZRFC, 2014, S. 14-17.

⁶⁰ Vgl. Nigrini, Mark J., The detection of income evasion through an analysis of digital distributions, Dissertation, Department of Accounting, Univ. Cincinnati, Cincinnati OH. und Nigrini, Mark J., A Taxpayer Compliance Application of Benford's Law, in: The Journal of the American Taxation Association, 1996, S. 72-91.

⁶¹ Vgl. Posch, Peter N., Ziffernanalyse in der Fälschungsaufspürung. Benford's Gesetz und Steuererklärungen in Theorie und Praxis, http://posch.org/paper/posch_faelschungenbenford.pdf (Abfrage vom 28.03.2015), S. 16-34.

⁶² Dieser wird detailliert in Abschnitt 3.4.1 erläutert.

⁶³ Vgl. Watrin, Christoph/ Struffert, Ralf, Benford's Law und Chi-Quadrat-Test – Chancen und Risiken des Einsatzes bei steuerlichen Prüfungen, in: Der Betrieb, 2006, S. 1748.

⁶⁴ Beschluss des Bundesgerichtshofes, 4 StR 419/89 vom 14.12.1989, abgedruckt in: Neue Juristische Wochenschrift, 1990, S. 1549.

⁶⁵ FG Münster 8 V 5774/02 vom 05.12.2002, sowie FG Münster 6 V 4562/03 E, U vom 10.11.2003, EFG 2004, 236.

⁶⁶ FG Münster 8 V 2651/03 E,U vom 14.08.2003, openJur 2011, 26048.

2.2 Analyse von makroökonomischen und statistischen Daten

Zur Untersuchung makroökonomischer Daten fand der Benford-Test erstmals im Jahr 2007 durch *Nye* und *Moul* Anwendung. Zunächst bestätigten sie die vorhandene Benford-Verteilung für Daten des Bruttoinlandsproduktes, welche sie mittels einer Monte-Carlo-Simulation generierten. Im Anschluss betrachteten sie in einer empirischen Untersuchung die Daten 183 Länder, wobei im Falle dieser zwischen OECD-Staaten und afrikanischen Ländern unterschieden wurde. Hierbei wiesen die Daten der OECD-Staaten eine deutlich bessere Datenqualität auf. Als Resümee empfehlen sie die Anwendung des Benford-Testes zur Analyse makroökonomischer Daten unter der Einschränkung, dass eine Bestätigung mittels weiterer Verfahren erfolgen sollte.⁶⁷

Gonzalia-Garcia und *Pastor* sprechen sich hingegen gegen die Nutzung der Ziffernanalyse zur Beurteilung der Qualität makroökonomischer Daten aus. Basis ihrer Untersuchung waren hierbei Daten aus dem Bankensektor sowie Zahlungsbilanzen der nationalen VGR und nationaler Finanzstatistiken von 80 verschiedenen Ländern der Jahre 1945 bis 2007. Mittels statistischer Testverfahren ermittelten sie die Deviation von der Benford-Verteilung der einzelnen Länder und verglichen diese mit dem Platz innerhalb eines bereits erstellten Rankings bzgl. der Datenqualität. Da sie hierbei keine Korrelation feststellen konnten, fassten sie dies als Indiz gegen die Wirksamkeit des Benford-Testes auf.⁶⁸

Den Beweis für die Eignung zur effizienten Manipulationsaufdeckung lieferten jedoch im Jahr 2011 *Rauch et al.* Sie analysierten makroökonomische Daten der damaligen 27 EU-Staaten und bildeten ein Ranking bzgl. der Abweichung von der Benford-Verteilung. Als Ergebnis wies ausgerechnet Griechenland, das bereits im Jahr 2003 durch das Statistische Amt der Europäischen Union der Täuschung überführt worden war, die größte Abweichung auf. Dies verdeutlicht, dass der Benford-Test generell ein geeignetes Mittel zur Beurteilung der Datenqualität darstellt.⁶⁹ Ergänzt wurde diese Untersuchung durch *Rauch et al.* in einer weiteren Analyse, in welcher ausschließlich statistische Daten aus dem sozialen Bereich betrachtet wurden. Bei diesen liegt grundsätzlich keine oder eine nur äußerst geringe Motivation zur Manipulation vor.

⁶⁷ Vgl. Nye, John/ Moul, Charles, The Political Economy of Numbers: On the Application of Benford's Law to International Macroeconomic Statistics, in: The B.E. Journal of Macroeconomics, 2007, S. 5-10.

⁶⁸ Vgl. Gonzalez-Garcia, Jesus/ Pastor, Gonzalo, Benford's Law and Macroeconomic Data Quality, in: International Monetary Fund, Working Paper, 2009, S. 7-10.

⁶⁹ Vgl. Rauch, Bernhard/ Götttsche, Max/ Brähler, Gernot/ Engel, Stefan, Fact and Fiction in EU-Governmental Economic Data, in: German Economic Review, 2011, S. 253.

Insgesamt war die Datenqualität nahezu aller Staaten deutlich besser als in der vorherigen Untersuchung. Griechenland wies hier sogar die zweitbeste Annäherung an die Benford-Verteilung auf, was ebenfalls klar für die Anwendbarkeit der Ziffernanalyse zur Entdeckung von Falschangaben spricht.⁷⁰

Eine weitere Untersuchung von *Rauch et al.* im Themenbereich der makroökonomischen Daten widmete sich den EU-Aufnahmekandidaten. Grundsätzlich liegt bei diesen die Motivation vor, eigene Daten zu fälschen, um den Beitrittsprozess in die Europäische Union bzw. in die Eurozone zu beschleunigen. Auffällig hohe Deviationen stellten sie bei den Ländern Island, Bosnien-Herzegowina und Serbien fest. Sie empfehlen eingehend detailliertere Prüfungen durchzuführen, ehe die Aufnahme in die EU erfolgt, um eine Eskalation wie im Falle Griechenlands frühzeitig zu unterbinden.⁷¹

Michalski und *Stoltz* untersuchten Zahlungsbilanzen mit Hilfe des Benfordschen Gesetzes und zeigten, dass bei Ländern mit Währungen gekoppelter Wechselkurse im Falle stark negativer Netto-Auslandsvermögen und negativer Leistungsbilanzen eine auffällige Abweichung von der Benford-Verteilung aufweisen.⁷²

Im Anwendungsfeld der statistischen Daten prüfte *Diekmann* zunächst im Jahr 2007, welche statistischen Größen grundsätzlich konform mit der Benford-Verteilung sind. So betrachtete er je 1.000 Mittelwerte, Varianzen, Korrelationskoeffizienten sowie standardisierte und nicht standardisierte Regressionskoeffizienten des *American Journal of Sociology* der Jahre 1996 und 1997. Von den untersuchten Größen wiesen jedoch lediglich die nicht-standardisierten Regressionskoeffizienten eine Benford-Verteilung auf.⁷³

Die Forscher *Günnel* und *Tödter* zogen Regressionskoeffizienten und Standardfehler der Zeitschriften *Empirica* und *Applied Economic Letters* heran. Sie verglichen die Benford-Deviation

⁷⁰ Vgl. Rauch, Bernhard/ Götsche, Max/ Brähler, Gernot/ Kronfeld, Thomas, Deficit versus social statistics: empirical evidence for the effectiveness of Benford's law, in: Applied Economic Letters, 2014, Vol. 21, No. 3, S. 147-151.

⁷¹ Vgl. Rauch, Bernhard/ Götsche, Max/ Brähler, Gernot/ Brinkert, Daniel/ Kronfeld, Thomas, European Integration and Statistical Credibility: Application of Benford's Law on Governmental Economic Data of EU Membership Aspirants, in: The Empirical Economics Letters, 2014, Vol. 13, No. 7, S. 755 – 762.

⁷² Michalski T./ Stoltz G., Do countries falsify economic data strategically? Some evidence that they might, in: Review of Economics and Statistics, Vol. 94, S. 2.

⁷³ Vgl. Diekmann, Andreas, Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data, in: Journal of Applied Statistics, 2007, S. 323, 324.

mit derjenigen von Daten des *Consensus Forecast*⁷⁴. Die Regressionskoeffizienten und Standardfehler zeigten eine gute Annäherung an die Benford-Verteilung, welche jedoch im Falle der Prognosedaten abgelehnt werden musste. Zu begründen sei dies mit häufigen Rundungen, welche bei Erstellung der Daten vorgenommen werden.⁷⁵ Um die Umstände genauer zu ergründen, betrachtete Tödter die Einzeldaten der in der obigen Untersuchung herangezogenen Artikel. Bei den ersten Ziffern trat jedoch in 25% der Fälle eine signifikante Abweichung auf. Er begründet dies mit der besonders großen Auswirkung der Manipulation der ersten Ziffern. Somit überwiege die Motivation für eine Korrektur zugunsten der gewünschten Ergebnisse in diesem Fall.⁷⁶ Diekmann und Jann hielten Tödter hingegen vor, er habe nicht geprüft, ob nicht-manipulierte Regressionskoeffizienten einer Benford-Verteilung tatsächlich folgen. Somit kann nicht beurteilt werden, ob die Ursache der Abweichungen in gezielten Manipulationen der Daten zu finden ist.⁷⁷

Auch Bauer und Gross kommen zu dem Schluss, dass die Ziffernanalyse eine grundsätzliche Eignung zur Manipulationsaufdeckung bei statistischen Daten besitzt. Im Falle kleiner Abweichungen müsste der Datenbestand jedoch entsprechend groß sein, damit die Täuschungen aufgedeckt werden können.⁷⁸

2.3 Sonstige Anwendungen

Zusätzlich zu den bereits dargestellten Themengebieten der Rechnungslegung, den makroökonomischen Daten sowie der Statistik verfügt das Benford's Law über eine Vielzahl weiterer Themengebiete. Im Folgenden wird ein Auszug wesentlicher Beispiele gegeben.

Die beiden Forscher Ley und Varian nutzten die Ziffernanalyse dazu, psychologische Preisbarrieren in dem Aktien-Index „Dow-Jones“ zu identifizieren.⁷⁹ In einer weiteren Untersuchung

⁷⁴ Der Consensus Forecast bildet auf Basis von Umfragedaten führender Ökonomen Schätzwerte und Prognosedaten über die zukünftige Entwicklung der Konjunktur.

⁷⁵ Vgl. Günnel, Stefan/ Tödter, Karl-Heinz, Does Benford's Law hold in economic research and forecasting?, in: Empirica, 2009, S. 279-289.

⁷⁶ Vgl. Tödter, Karl-Heinz, Benford's Law as an Indicator of Fraud in Economics, in: German Economic Review, 2009, S. 343-346.

⁷⁷ Vgl. Diekmann, Andreas/ Jann, Ben, Benford's Law and Fraud Detection. Facts and Legends, in: German Economic Review, Vol. 11, No. 3, 2010, S. 397 – 401.

⁷⁸ Vgl. Bauer, Johannes/ Gross, Jochen, Difficulties in Detecting Fraud? The Use of Benford's Law on Regression Tables, in: Journal of Economics and Statistics, 2011, S. 744-748.

⁷⁹ Vgl. Ley, Eduardo/ Varian, Hal R., Are there psychological barriers in the Dow-Jones index?, in: Applied Financial Economics, 1994, S. 223.

bewies *Ley* die Benford-Konformität für Tages-Renditen des Dow-Jones-Indexes.⁸⁰ *Rodriguez* hingegen widersprach dieser These in seiner Forschungsarbeit. Laut seinen Resultaten sind die Schlusskurse des Dow-Jones-Indexes nicht Benford-verteilt.⁸¹ Gemäß *Tödter* folgen die Schlusskurse des deutschen DAX hingegen sehr wohl dem Benfordschen Gesetz.⁸² Eine weitere Forschergruppe um *De Ceuster et al.* unternahm ebenfalls den Versuch, psychologische Preisbarrieren von Aktien-Indizes mittels der Ziffernanalyse aufzuspüren. Im Gegensatz zu den vorherigen Untersuchungen zeigten sie jedoch zunächst, dass im Falle der Aktienkurse eine Benford-Verteilung vorliegt. Im Anschluss führten sie die Analyse auf Preisbarrieren durch, dessen Resultate jedoch keine Bestätigung solcher hervorbrachten. Zu begründen sei dies damit, dass die Anfangsziffern der Schlusskurse eine Benford-Verteilung aufweisen und keine Gleichverteilung, wie *Ley* und *Varian* annahmen. Dies sei grundsätzlich ursächlich für die falsch-positiven Ergebnisse der beiden Wissenschaftler.⁸³

Ein weiteres Applikationsfeld des Benford's Law umfasst die Untersuchung von Auktionspreisen des Internet-Auktionsanbieters *eBay*. Der Wissenschaftler *Giles* wollte mittels des Benford-Testes nachweisen, dass Preismanipulationen aufgrund einer persönlichen Verbindung zwischen Verkäufer und den Bietern vorgenommen werden, um die Preise künstlich zu erhöhen. Da die Auktionspreise jedoch insgesamt Benford-verteilt sind, konnte er diese These nicht bestätigen.⁸⁴

Auch in Bereichen der Politik wurde die Ziffernanalyse unter Verwendung der Benford-Verteilung zur Aufdeckung von Manipulationen angewendet. So wendeten *Cho* und *Gaines* den Benford-Test auf Finanzdaten der Parteien im US-Wahlkampf an und ermittelten deutliche Abweichungen von der Benford-Verteilung. Als Resümee empfehlen sie die grundsätzliche Anwendung der Ziffernanalyse in diesem Themenfeld und sprechen sich eindeutig für eine tiefergehende Analyse der Daten aus.⁸⁵

⁸⁰ Vgl. *Ley*, Eduardo, On the Peculiar Distribution of the U.S. Stock Indexes' Data, in: *The American Statistician*, 1996, S. 312.

⁸¹ Vgl. *Rodriguez*, Ricardo J., First Significant Digit Patterns from Mixtures of Uniform Distributions, in: *The American Statistician*, 2004, S. 68.

⁸² Vgl. *Tödter*, Karl-Heinz, Das Benford-Gesetz und die Anfangsziffern von Aktienkursen, in: *WiSt*, 2007, S. 94.

⁸³ Vgl. *De Ceuster*, Marc J. K./ *Dhaene*, Geert/ *Schateman*, Tom, On the hypothesis of psychological barriers in stock markets and Benford's Law, in: *Journal of Empirical Finance*, 1998, S. 277-279.

⁸⁴ Vgl. *Giles*, David E., Benford's Law and naturally occurring prices in certain eBay auctions, in: *Applied Economic Letters*, 2007, S. 159, 160.

⁸⁵ Vgl. *Cho*, Wendy K. Tam/ *Gaines*, Brian J., Breaking the (Benford) Law: Statistical Fraud Detection in Campaign Finance, in: *The American Statistician*, 2007, S. 221-223.

Zur Aufdeckung von Wahlbetrügereien empfiehlt *Mebane* den Einsatz des Benford-Testes,⁸⁶ was jedoch von *Deckert et al.* dementiert wurde.⁸⁷ Mittels der Ziffernanalyse gelang es *Beber* und *Sacco* hingegen, Wahlbetrügereien in den umstrittenen Präsidentschaftswahlen in Senegal (2007) und Nigeria (2003) aufzudecken. Hierbei unterstellten sie im Falle der letzten Ziffern sowie dem Paar der letzten beiden Ziffern eine Gleichverteilung.⁸⁸ *Breunig* und *Goerres* wendeten auf diesen Erkenntnissen aufbauend das Benfordsche Gesetz auf die zweiten Ziffern der abgegebenen Stimmen auf Wahlbezirksebene bei Wahlen des Deutschen Bundestags zwischen 1990 und 2005 an, konnten jedoch keine gravierenden Betrügereien aufdecken.⁸⁹

Rauch et al. widmeten sich in einer weiteren Forschungsarbeit den Rechenschaftsberichten deutscher Parteien. Zum einen konnten sie eine auffällig hohe Benford-Deviation bei den „Grauen Panther“ nachweisen, welche bereits in der Vergangenheit der Manipulationen überführt wurden. Somit konnten sie zeigen, dass der Benford-Test durchaus eine Berechtigung zur Anwendung im Rahmen parteipolitischer Finanzdaten aufweist. Zum anderen wiesen kleinere Parteien i.A. eine geringe Benford-Konformität auf. Dies begründeten sie mit möglicherweise fingierten privaten Parteispenden, da deren Zunahme mit einer erhöhten staatlichen Teilfinanzierung einhergeht, was v.a. für die kleinen Parteien als äußerst lukrativ erscheint. Aus diesem Grund sollten speziell die Rechenschaftsberichte dieser genausten Untersuchungen unterzogen werden.⁹⁰

Zur Aufdeckung von Manipulationsversuchen in Interviews nutzten *Schäfer et al.* die Ziffernanalyse unter Verwendung des Benford's Law.⁹¹ *Judge* und *Schechter* gelang es mittels des Benford-Testes als Ursachen der Abweichungen Zählfehler sowie falsch gestellte Fragen zu

⁸⁶ Vgl. *Mebane, W.*, Election forensics: The second-digit Benford's law test and recent American presidential elections, in: *Election Fraud: Detecting and Deterring Electoral Manipulation*, edited by *R. Alvarez, T. Hall und S. Hyde*, Washington DC, 2008, S. 161-181 und *Mebane, W.*, Fraud in the 2009 presential election in Iran?, in: *Chance*, 2010, S. 6-15 sowie *Mebane, W.*, Comment on "Benford's Law and the Detection of Election Fraud", in: *Political Analysis*, 2011, S. 269-272.

⁸⁷ Vgl. *Deckert, J./ Myagkov, M./ Ordeshook, P.*, Benford's Law and the Detection of Election Fraud, in: *Political Analysis*, 2011, S. 245-268.

⁸⁸ Vgl. *Beber, B./ Scacco, A.*, What the Numbers Say: A Digit-Based Test for Election Fraud, in: *Political Analysis*, 20. Jg., Nr. 2, S. 211-234.

⁸⁹ Vgl. *Breunig, C./ Goerres, A.*, Searching for electoral irregularities in an established democracy: Applying Benford's Law tests to Bundestag elections in Unified Germany, in: *Electoral Studies*, 2011, 30. Jg., Nr. 3, S. 534-545.

⁹⁰ Vgl. *Rauch, Bernhard/ Götsche, Max/ Brähler, Gernot/ Geidel, Felix/ Pietras, Thomas*, Überprüfung der Rechenschaftsberichte deutscher Parteien mit Hilfe des Benford's Law, in: *BFuP*, 2014, S. 175-191.

⁹¹ Vgl. *Schäfer, Christin/ Schröpler, Jörg-Peter/ Müller, Klaus-Robert/ Wagner, Gert G.*, Automatic Identification of Faked and Fraudulent Interviews in the German SOEP, in: *Schmollers Jahrbuch*, 2005, S. 186-188.

identifizieren. Als Grundlage dienten ihnen Umfragedaten aus der Agrarbranche.⁹² *Schräpler* konnte mittels des Benford's Law Manipulationen im „Socio-Economic-Panel“ (SOEP) aufdecken.⁹³

Die Staatsausgaben für das Militär verschiedener Länder untersuchten *Rauch et al.* mit der Ziffernanalyse. Sie stellten v.a. bei den USA eine auffällig hohe Abweichung fest, welche in eine Vielzahl militärischer Interaktionen involviert sind bzw. waren und somit eine besonders große Motivation zur Anpassung eigener Daten haben, um Rückschlüsse aus den Finanzdaten auf die militärischen Bestände zu verhindern.⁹⁴

Stephan Kienle bewies in einer Veröffentlichung, dass auch Deckungsbeiträge von in Deutschland herausgegebenen Pfandbriefen eine Benford-Konformität aufweisen. Er setzte hierbei auf den Chi²-Anpassungstest, der Mean-Absolute-Deviation und den Mittelwert-Test.⁹⁵

⁹² Vgl. Judge, George/ Schechter, Laura, Detecting Problems in Survey Data Using Benford's Law, in: The Journal of Human Resources, 2009, S. 1-24.

⁹³ Vgl. Schräpler, J.-P., Benford's Law as an Instrument for Fraud Detection in Surveys Using the Data of the Socio-Economic Panel (SOEP), in: Jahrbücher für Nationalökonomie und Statistik, 231. Jg., Nr. 5-6, S. 212-234.

⁹⁴ Vgl. Rauch, Bernhard/ Götsche, Max/ Langenegger, Stephan, Detecting Problems in Military Expenditure Data Using Digital Analysis, in: Defence and Peace Economics, 2014, Vol. 25, No. 2, S. 97-111.

⁹⁵ Vgl. Kienle, Stephan, What Benford Can Tell Us About Cover Pools – An Empirical Analysis, in: International Business & Economics Research Journal, Vol. 14, No. 6, 2015, S. 829.

3. Grundlagen des Benford's Law

3.1 Beschreibung der Gesetzmäßigkeit

Der Ursprung des Gesetzes findet sich im Jahr 1881, in welchem der Astronom *Simon Newcomb* eine interessante Entdeckung verzeichnete: Bei der Betrachtung von Logarithmentafeln stellte er fest, dass die ersten Seiten mit niedrigen Anfangsziffern deutlich abgenutzter erschienen, als die hinteren der großen Zahlen. Hieraus schloss er, dass kleinere Zahlen deutlich häufiger nachgeschlagen würden als große und folgerte weiter, dass es einen systematischen Zusammenhang zwischen der Anfangsziffern und ihrer Häufigkeiten geben muss.⁹⁶ Doch erst mehr als 50 Jahre später im Jahr 1931 gewann die Erkenntnis Newcombs in der Wissenschaft größere Bedeutung. *Frank Benford*, der unabhängig von *Newcomb* auf den gleichen Sachverhalt stieß, zeigte für eine Vielzahl an Zahlen natürlichen Ursprungs wie z.B. die Länge von Flüssen, die Höhe von Bergen, Stromrechnungen und Baseballergebnissen, dass die Anfangsziffern nicht gleichverteilt sind, sondern einer logarithmischen Verteilung folgen.⁹⁷ Zunächst wurde *Benford* vorgeworfen, eigene Daten gefälscht zu haben,⁹⁸ in einer weiteren Untersuchung durch *Hill* konnte dieser jedoch zeigen, dass das Auftreten dieser Verteilung, welche auch als Benford-Verteilung bezeichnet wird, auch ohne Manipulationen zu beobachten ist.⁹⁹ Für die erste Ziffer gilt die folgende Formel zur Berechnung der Wahrscheinlichkeit:

$$P(D_1 = d_1) = \log_{10} \left(1 + \frac{1}{d_1} \right). \quad (1)$$

Hierbei bezeichnet $P(D_1)$ die Wahrscheinlichkeit für das Auftreten der Zufallsvariablen D_1 der ersten Ziffer und d_1 für die tatsächlich eingesetzte Ziffer. Für diese gilt $d_1 \in (0; 9]$ und $d_1 \in \mathbb{Z}$. d_1 kann also die ganzzahligen Werte zwischen 1 und 9 annehmen. Grafisch ist die Verteilung der ersten Ziffer in Abbildung 1 dargestellt.

⁹⁶ Vgl. Newcomb, Simon, Note on the Frequency of Use of the Different Digits in Natural Numbers, in: American Journal of Mathematics, Vol. 4, No. 1, 1881, S. 39f.

⁹⁷ Vgl. Benford, Frank, The Law of Anomalous Numbers, in: Proceedings of the American Philosophical Society, 1938, S. 553.

⁹⁸ Vgl. Diaconis, Persi/ Freedman, David, On Rounding Percentages, in: Journal of the American Statistical Association, Vol. 74, No. 366, 1979, S. 359-364.

⁹⁹ Vgl. Hill, Theodore P., A statistical derivation of the significant digit law, in: Statistical Science, Vol. 10, No. 4, 1995, S. 355.

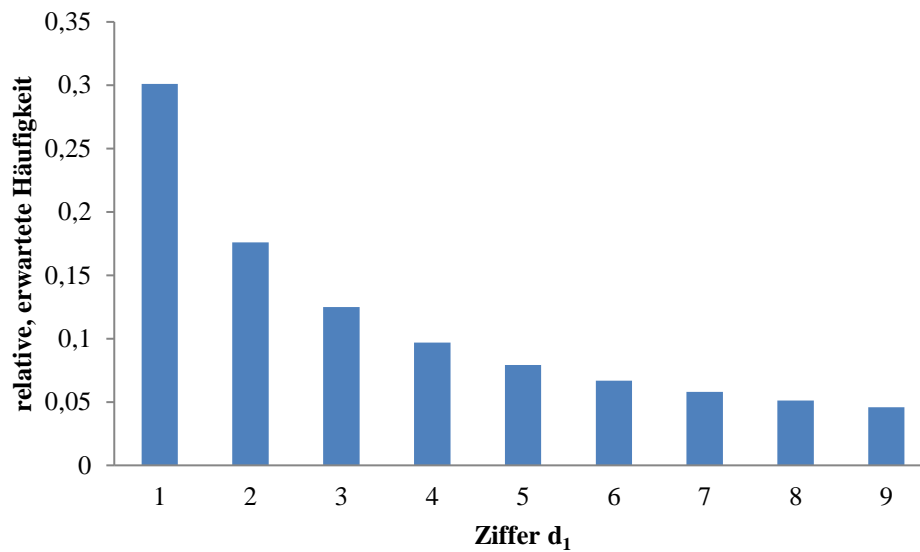


Abbildung 1: Verteilung der ersten Ziffer d_1 des Benfordschen Gesetzes

Für die zweite Ziffer gilt die Berechnungsformel:

$$P(D_2 = d_2) = \sum_{d_1=1}^9 \log_{10} \left(1 + \frac{1}{10 \cdot d_1 + d_2} \right). \quad (2)$$

Für d_2 gilt hierbei $d_2 \in [0; 9]$ und $d_2 \in \mathbb{Z}$. Im Vergleich zur ersten Ziffer kann d_2 zusätzlich den Wert 0 annehmen. Als allgemein formuliertes Gesetz gilt für die Verteilung der Stelle k einer Zahl die folgende Gleichung:

$$P(D_1 = d_1, \dots, D_k = d_k) = \log_{10} \left[1 + \frac{1}{\sum_{i=1}^k d_i \cdot 10^{k-i}} \right]. \quad (3)$$

Hierbei gilt entsprechend $k \in \mathbb{Z}$, $d_1 \in (0; 9]$ und $d_j \in [0; 9]$ mit $j = 2, \dots, k$ sowie $d_1, d_j \in \mathbb{Z}$. Diese verallgemeinerte Formel wird in der Literatur auch häufig als „General Significant Digit Law“ bezeichnet.¹⁰⁰ Tabelle 1 gibt einen Überblick über die Wahrscheinlichkeiten des Auftretens der ersten vier Stellen einer Zahl. Betrachtet man die Wahrscheinlichkeiten der vierten Stelle so fällt auf, dass hier nahezu alle Ziffern mit der gleichen Häufigkeit vorkommen. Allgemein formuliert konvergiert die Benford-Verteilung für höhere Stellen gegen eine Gleichverteilung. Praktisch kann diese nach der vierten Ziffer unterstellt werden.¹⁰¹

¹⁰⁰ Vgl. Hill, Theodore P., A statistical derivation of the significant digit law, in: Statistical Science, Vol. 10, No. 4, 1995, S. 354.

¹⁰¹ Vgl. Posch, Peter N., Ziffernanalyse in Theorie und Praxis – Testverfahren zur Fälschungsaufspürung mit Benfords Gesetz, Aachen 2005, S.6.

d_i	$P(d_1)$	$P(d_2)$	$P(d_3)$	$P(d_4)$
0		0,11968	0,10178	0,10018
1	0,30103	0,11389	0,10138	0,10014
2	0,17609	0,10882	0,10097	0,10010
3	0,12494	0,10433	0,10057	0,10006
4	0,09691	0,10031	0,10018	0,01002
5	0,07918	0,09668	0,09979	0,09998
6	0,06695	0,09337	0,09940	0,09994
7	0,05799	0,09035	0,09902	0,09990
8	0,05115	0,08757	0,09864	0,09986
9	0,04576	0,08500	0,09827	0,09982

Tabelle 1: Wahrscheinlichkeiten des Auftretens der ersten bis vierten Ziffer der Benford-Verteilung

Quelle: Diekmann, Andreas, Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data, in: Journal of Applied Statistics, Vol. 34, No. 3, 2007, S. 323.

Im Rahmen der Berechnung der Wahrscheinlichkeiten wird angenommen, dass die Zahl in der sog. „Zehnerpotenz-Schreibweise“ vorliegt bzw. in diese konvertiert wird. Demnach bildet die erste Ziffer die erste Stelle und liegt auf dem Intervall $[1,10)$. Konkret bedeutet dies, dass z.B. die Zahl 795 in Zehnerpotenz-Schreibweise $7,95 \cdot 10^2$ entspricht. Somit besteht eine eindeutige Unabhängigkeit zwischen dem Auftreten einer Ziffer und der Zehnerpotenz einer Zahl. Relevant ist lediglich ihre Mantisse. Die Zahlen $3,97 = 3,97 \cdot 10^0$ und $3.970.000 = 3,97 \cdot 10^6$ weisen in der Folge identische Wahrscheinlichkeiten auf. Für beide gilt $P(D_1 = 3, D_2 = 9, D_3 = 7) = \log_{10} \left(1 + \frac{1}{397} \right) = 0,00109257$.

3.2 Voraussetzungen und Eigenschaften

Die Basis für das Vorliegen einer Benford-Verteilung ist die Tatsache, dass die Mantissen der logarithmierten Daten gleichverteilt zwischen 0 und 1 sind. Dies ist häufig dann der Fall, wenn die Zahlen einem exponentiellen Wachstumspfad folgen.¹⁰² Damit dies der Fall ist, muss der entsprechende Datenbestand einige Voraussetzungen erfüllen.

¹⁰² Vgl. Nigrini, Mark J., Digital Analysis Using Benford's Law: Test & Statistics for Auditors, 2. Aufl., Vancouver, 2000, S.10.

In der Regel ist es nicht möglich, die gesamte Grundgesamtheit in eine Untersuchung einfließen zu lassen, sodass man meist auf die Erhebung von Stichproben angewiesen ist. Mit steigendem Stichprobenumfang ist hierbei eine höhere Konformität mit der Benford-Verteilung zu beobachten.¹⁰³ Die Stichproben müssen die Voraussetzung erfüllen, dass sie zufälligen Ursprungs sind, d.h. die Daten dürfen keinesfalls systematisch ausgewählt worden sein. Insbesondere ist deshalb auf Auswahlverfahren zu verzichten, da diese sämtliche Analysen unbrauchbar machen würden. Zudem dürfen Daten nicht zugewiesen worden sein, wie es beispielsweise bei Adressdaten, Kontonummern oder Telefonnummern der Fall ist.¹⁰⁴

Auch sollten die Einheiten der Daten die gleiche Dimension aufweisen, damit eine Vergleichbarkeit gewährleistet werden kann. Demzufolge ist es generell möglich, beispielsweise Daten der Einheit Euro und US-Dollar miteinander zu vergleichen, nicht jedoch Kilometer und Kilogramm. Auch sollte der Datensatz frei von fixen Ober- bzw. Untergrenzen sein und keine Schwellenwerte auftreten, da diese allesamt zu einer Verzerrung der Häufigkeitsverteilung führen und somit falsch-positive Resultate hervorgebracht würden.¹⁰⁵

Tendenziell begünstigt die Tatsache, dass sich mehr kleine als große Zahlen in einem Datensatz befinden, das Vorhandensein einer Benford-Verteilung. Aus diesem Grund sollte der Median der Daten kleiner sein als deren Mittelwert, sowie die Schiefe größer Null, die Verteilung also rechtsschief sein.¹⁰⁶

Statistische Kennzahlen wie Mittelwerte, Quantile und Standardabweichungen sind i.d.R. nicht Benford-verteilt. Der Grund dafür liegt in der Tatsache, dass diese aus den Rohdaten abgeleitet werden und somit eigenen statistischen Verteilungen folgen. Aus diesem Grund sollten in dem zu untersuchenden Datensatz keine solche Werte enthalten sein.¹⁰⁷

¹⁰³ Vgl. im Folgenden: Mochty, Ludwig, Die Aufdeckung von Manipulationen im Rechnungswesen – Was leistet Benford's Law?, in: Die Wirtschaftsprüfung, 2002, S. 734 sowie Nagasaka, Kenji, On Benford's Law, in: Annals of the Institute of Statistical Mathematics, 1984, S. 337.

¹⁰⁴ Vgl. Quick, Reiner/ Wolz, Matthias, Benford's Law in deutschen Rechnungslegungsdaten, in: BFuP, 2003, S. 213.

¹⁰⁵ Vgl. Nigrini, Mark J., Digital Analysis Using Benford's Law: Test & Statistics for Auditors, 2. Aufl., Vancouver 2000, S. 24.

¹⁰⁶ Vgl. Nigrini, Mark J., Digital Analysis Using Benford's Law: Test & Statistics for Auditors, 2. Aufl., Vancouver 2000, S. 38.

¹⁰⁷ Vgl. Leemis, Lawrence M./ Schmeiser, Bruce W./ Evans, Diane L., Survival Distributions Satisfying Benford's Law, in: American Statistician, 2000, S. 236ff.

Die Benford-Mengen weisen i.d.R. wichtige mathematische Eigenschaften auf, welche im Rahmen von Analysen interessante Vorteile mit sich bringen. Bei diesen handelt es sich um die Invarianz ggü. mathematischer Operationen, die Skaleninvarianz und die Baseninvarianz.

Die Invarianz ggü. mathematischer Operationen besagt, dass Benfordmengen beliebig miteinander addiert, subtrahiert, multipliziert oder dividiert werden dürfen, ohne dass die vorhandene Benford-Verteilung verloren geht. Zu beachten ist hierbei lediglich, dass Werte, welche durch mathematische Operationen berechnet wurden, nicht doppelt erfasst werden. Dies bedeutet insbesondere, dass Operanden und deren Ergebnisse nicht gleichzeitig erhoben werden dürfen. Eine doppelte Erhebung würde ebenfalls zu Verzerrungen führen.¹⁰⁸

Die Skalenvarianz beschreibt die Eigenschaft, dass das Skalenniveau beliebig veränderbar ist. Dies bedeutet, dass ein Benford-verteilter Datensatz problemlos mit einer Konstante ungleich Null multipliziert oder dividiert werden kann, ohne dass dies die Verteilung beeinflusst.¹⁰⁹ Somit sind insbesondere Umrechnungen von Einheiten (z.B. Pfund in Kilogramm oder Euro in US-Dollar) möglich.

Die Baseninvarianz bezieht sich auf die Zahlenbasis der Benfordmenge. So ist das Benfordsche Gesetz zum einen für sämtliche Basen wie z.B. das Binärsystem, Hexadezimalsystem oder das Oktalsystem gültig. Zum anderen sind auch Umrechnungen zwischen den Systemen möglich. Zu beachten ist hierbei lediglich, dass sich die Anzahl der Ziffern sowie deren Verteilungen in Abhängigkeit der Zahlenbasis ändern. Im Falle des Binärsystems treten somit nur die Ziffern null und eins, im Falle des Oktalsystems die Ziffern eins bis acht auf.¹¹⁰

3.3 Eignung als Indikator von Manipulationen

Die grundlegend unterstellte Annahme, möchte man mit Hilfe der Ziffernanalyse unter Verwendung des Benfordschen Gesetzes Manipulationen in Datensätzen aufdecken, ist die, dass

¹⁰⁸ Vgl. Boyle, Jeff, An Application of Fourier Series to the most Significant Digit Problem, in: The American Mathematical Monthly, 1994, S. 879, 883 sowie Schatte, Peter, On mantissa distributions in computing and Benford's Law, in: Journal of information processing and cybernetics, 1988, S. 443-455.

¹⁰⁹ Vgl. Pinkham, Roger S., On the Distribution of First Significant Digits, in: The Annals of Mathematical Statistics, Vol. 32, No. 4, 1961, S. 1223-1230.

¹¹⁰ Vgl. Hill, Theodore P., Base-Invariance Implies Benford's Law, in: Proceedings of the American Mathematical Society, Vol. 123, No. 3, 1995, S. 888, 889.

unmanipulierte Daten eine Übereinstimmung mit der Benford-Verteilung aufweisen, diese jedoch durch die vorgenommenen Falsch- und Fehlangaben verloren geht. Mit steigender Anzahl manipulierter Daten würde demnach auch die Abweichung von der Benford-Verteilung zunehmen. Aus diesem Grund gilt es mittels verschiedener statistischer Testverfahren und Kennzahlen, welche in Abschnitt 3.4 näher erläutert werden, die Benford-Deviation zu bestimmen, um somit Rückschlüsse auf mögliche Manipulationen zu ziehen.¹¹¹

Theoretisch ist es jedoch möglich, Manipulationen so durchzuführen, dass keine Abweichungen von der Benford-Verteilung entstehen. Grundsätzlich erscheint ein solches Vorgehen v.a. bei großen Datenmengen, aufgrund der äußerst komplexen Anforderungen, in der Praxis als zu aufwendig und damit sehr unwahrscheinlich.¹¹² Hill führte bezüglich dieser Problematik eine Untersuchung durch. Er bat randomisiert ausgewählte Probanden, ihnen zufällig in den Sinn kommende Zahlen zu notieren und prüfte diese Zahlen im Anschluss auf eine Benford-Verteilung. Diese musste er jedoch ablehnen, da die Daten eher einer Gleichverteilung folgten. Die genaue Verteilung im Vergleich zu der Benford- und der Gleichverteilung ist Abbildung 2 zu entnehmen.

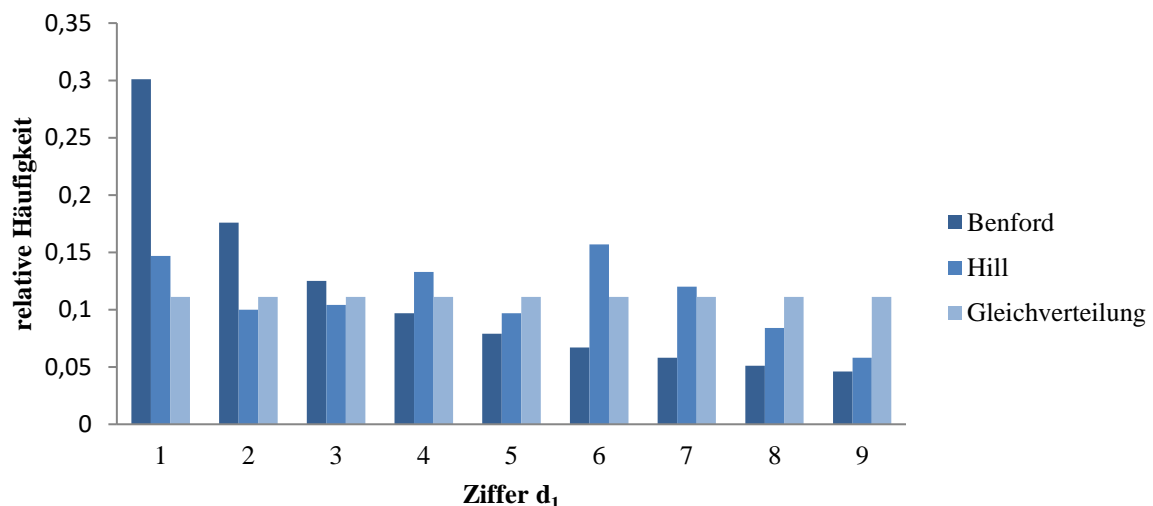


Abbildung 2: Vergleich der Verteilung der ersten Ziffern von Hill, Benford und der Gleichverteilung¹¹³

¹¹¹ Vgl. Watrin, Cristoph/ Struffert, Ralf, Benford's Law und Chi-Quadrat-Test – Chancen und Risiken des Einsatzes bei steuerlichen Prüfungen – Zugleich Anmerkungen zum Urteil des FG Münster vom 07.12.2005 1 K 6384/03 E, in: Der Betrieb, 2006, S. 1749.

¹¹² Vgl. Mochty, Ludwig, Die Aufdeckung von Manipulationen im Rechnungswesen – Was leistet Benford's Law?, in: Die Wirtschaftsprüfung, 2002, S. 729.

¹¹³ In Anlehnung an Hill, Theodore P., Random-Number guessing and the first digit phenomenon, in: Psychological Reports, 1988, S. 969.

Eine deutliche Einschränkung der o.g. Untersuchung war jedoch, dass die Probanden nicht darüber in Kenntnis gesetzt wurden, was *Hill* eigentlich untersuchen wollte. Demnach könnte man zu dem Schluss kommen, dass andere Ergebnisse erzielt worden wären, würde man die Manipulationen gezielt Benford-konform durchführen. Dieser Problematik gingen *Watrin et al.* in einer weiteren Studie nach. Diese unterteilte sich in zwei Stufen: Zunächst wurden die Probanden gebeten, Daten aus der Rechnungslegung eines fiktiven Unternehmens zu manipulieren. In einem zweiten Schritt wurden die Teilnehmer dann über den Untersuchungszweck aufgeklärt und gebeten, die Manipulationen so durchzuführen, dass weiterhin eine Benford-Verteilung zu beobachten ist. Im Ergebnisvergleich konnten sie keine signifikanten Unterschiede feststellen, was gegen Manipulationen in Kongruenz mit der Benford-Verteilung spricht.¹¹⁴

Klassische Prüfungshandlungen beschränken sich meist auf die Examinierung weniger großer Beträge, da hier die Erfolge im Falle der Aufdeckung von Fälschungen am größten sind. In der Praxis werden Manipulationen jedoch häufig auf eine Vielzahl kleinerer Beträge verteilt, welche somit unentdeckt bleiben. Der Benford-Test hingegen beschränkt sich nicht lediglich auf die Prüfung großer Beträge, sondern bezieht alle Zahlen in die Untersuchung ein, sodass die Ziffernanalyse hier den klassischen Prüfungshandlungen sogar überlegen ist.¹¹⁵

Doch auch die Nutzung des Benford-Testes ist mit einigen Einschränkungen verbunden. Zum ersten betrifft dies den Umfang der zu untersuchenden Stichprobe. Wird z.B. der Chi²-Anpassungstest zur Ermittlung der Benford-Deviation verwendet, der in Abschnitt 3.4.1 ausführlich erläutert wird, beeinflusst die Größe des Stichprobenumfanges die Testergebnisse signifikant: Ein großer Stichprobenumfang führt gemäß *Rodriguez* zu einem starken Anstieg der sog. „Alpha-Fehlern“^{116, 117}. Auch müssen die Daten die in Abschnitt 3.2 erläuterten Voraussetzungen erfüllen, andernfalls besitzen die Ergebnisse der Untersuchungen keinerlei Aussagekraft. Ein weiterer Nachteil ist, dass fehlende und doppelte Werte ebenfalls nicht zwingend erkannt werden.¹¹⁸

¹¹⁴ Vgl. Watrin, Christoph/ Struffert, Ralf/ Ullmann, Robert, Benford's Law: an instrument for selecting tax audit targets?, in: Review of Managerial Science, 2008, S. 235, 236.

¹¹⁵ Vgl. Baetge, Jörg/ Melcher, Thorsten/ Schulz, Roland, Vermeidung von Bilanzdelikten durch (Früh-) Erkennungsmethoden - Trends in der Wirtschaftsprüfung, in: George Herde (Hrsg.), 3. Deggendorfer Forum zur Digitalen Datenanalyse, Bilanzdelikte erkennen und vermeiden, Deggendorf 2008, S. 49.

¹¹⁶ Eine ausführliche Darstellung folgt in Abschnitt 3.4.1.

¹¹⁷ Vgl. Rodriguez, Ricardo J., Reducing False Alarms in the Detection of Human Influence on Data, in: Journal of Accounting, Auditing & Finance, 2004, S. 141-158.

¹¹⁸ Vgl. Durtschi, Cindy/ Hillison, William/ Pacini, Carl, The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data, Journal of Forensic Accounting, 2004, Jg. 5, Nr. 1, S. 17-34.

Vergleicht man die Vor- und Nachteile des Benford-Testes, so erscheint dieser ein sehr effizientes Instrumentarium zur Manipulationsaufdeckung zu sein, sofern die entsprechenden Voraussetzungen und Einschränkungen beachtet werden. Jedoch dürfen die Ergebnisse aufgrund der nicht vorhandenen Sicherheit nicht als Beweis, sondern lediglich als Indiz für mögliche Manipulationen aufgefasst werden, weshalb die verdächtigen Daten stets genaueren Untersuchungen unterzogen werden sollten. Hierbei wurde von *Odenthal* empfohlen, den Datensatz anhand struktureller Merkmale aufzugliedern, um auf Basis der entdeckten Abweichungen Schlussfolgerungen über die durchgeführten Manipulationen zu ziehen.¹¹⁹ Die Wissenschaftler *Busta* und *Weinberg* entsprachen den Empfehlungen *Odenthals* und konnten mit dieser Vorgehensweise innerhalb eines Datensatzes, der zu 10% manipulierte Daten enthielt, 64% der Falschangaben aufdecken.¹²⁰

3.4 Statistische Testverfahren und Maßzahlen

Wie bereits zuvor erläutert, lautet das Grundprinzip der Manipulationsaufdeckung mittels der Benford-Analyse, dass manipulierte und nicht-manipulierte Daten unterschiedliche Ziffernverteilungen aufweisen. Demnach wird in einem ersten Schritt zunächst eine Stichprobe erhoben, von welcher auf die zugehörige Grundgesamtheit geschlossen werden kann.¹²¹ Im Anschluss wird mittels diverser statistischer Methoden die Deviation der beobachteten Verteilung von der Benford-Verteilung bestimmt und geprüft, ob eine solche für die empirischen Daten angenommen werden kann.¹²² Im Folgenden werden verschiedene statistische Testverfahren und Maßzahlen detailliert erläutert, welche sich grundsätzlich zur Bestimmung des Abweichungsgrades von statistischen Verteilungen eignen.

¹¹⁹ Vgl. Odenthal, Roger, Digitale Ziffernanalyse: Ein wirkungsvoller Beitrag zur computergestützten Delikt Deliktrevision?, in: Die Wirtschaftsprüfung, 1999, S. 631, 633, 635.

¹²⁰ Vgl. Busta, Bruce/ Weinberg, Randy, Using Benford's law and neural networks as review procedure, in: Managerial Auditing Journal, 1998, S. 361-362.

¹²¹ Vgl. Watrin, Christoph/ Ullmann, Robert, Ziffernanalyse in der steuerlichen Betriebsprüfung: Voraussetzungen, Funktionsweise, Anwendungsmöglichkeiten, in: Die Wirtschaftsprüfung, 2009, S. 99.

¹²² Vgl. Bley Müller, Josef/ Gehlter, Günther/ Güblicher, Herbert, Statistik für Wirtschaftswissenschaftler, 15. Aufl., München 2008, S. 127.

3.4.1 Der Chi²-Anpassungstest

Einzuzuordnen ist der Chi²-Anpassungstest in die nicht-parametrischen Testverfahren der Statistik. Im Gegensatz zu den parametrischen Testverfahren werden hierbei keine Aussagen über einzelne Parameter gefällt, sondern der Frage nachgegangen, ob einzelne statistische Verteilungen empirische Daten zutreffend beschreiben oder aber abgelehnt werden müssen.¹²³

Hierzu wird in einem ersten Schritt wie bei allen statistischen Anpassungstests eine sog. „Nullhypothese“ – auch als H_0 bezeichnet – formuliert, wobei es das grundlegende Ziel des Chi²-Anpassungstestes ist, diese abzulehnen. Hierbei wird getestet, ob die unterstellte Verteilung (P_0) für den empirischen Datensatz zutreffend ist.¹²⁴ Grundsätzlich besteht bei diesen Signifikanztests stets die Gefahr des Auftretens sog. Fehler „erster Art“ und „zweiter Art“. Erstere, auch als „ α -Fehler“ bezeichnet, beschreiben die Tatsache, die Nullhypothese fälschlicherweise abzulehnen. Die Fehler zweiter Art oder auch „ β -Fehler“ treten auf, wenn die Nullhypothese angenommen würde, obwohl diese in Wahrheit nicht zutrifft.¹²⁵ Mittels der nichtparametrischen Testverfahren ist es jedoch lediglich möglich, eine der genannten Fehlerarten zu verringern, weshalb der Chi²-Anpassungstest ausschließlich die Fehler erster Art minimiert. Mit welcher Wahrscheinlichkeit dieser auftritt, wird im Vorfeld mittels des Signifikanzniveaus α angegeben.¹²⁶ Erfolgt dessen Festlegung auf 5%, so beträgt die Wahrscheinlichkeit für das Auftreten der α -Fehler 5%. Wird das Signifikanzniveau zu 1% definiert, sinkt die Irrtumswahrscheinlichkeit auf 1%. Somit ist die Chance höher, dass die getroffene signifikante Entscheidung korrekt ist. In der Literatur hat sich weitestgehend der Wert $\alpha = 0,05$ als praxistauglich durchgesetzt. Wie bereits beschrieben ist es jedoch keinesfalls möglich, Einfluss auf die Fehler zweiter Art zu nehmen.¹²⁷

Die ausschließliche Beeinflussung der Fehler erster Art führt zeitgleich dazu, dass die im Vorfeld definierte Nullhypothese lediglich abgelehnt, nicht jedoch angenommen werden kann. Entsprechend wird H_0 so formuliert, dass diese das Gegenteil von dem aussagt, was mittels der

¹²³ Vgl. Bückner, Rüdiger, Statistik für Wirtschaftswissenschaftler, 5. Aufl., Oldenburg, 2003, S. 207 und Bleymüller, Josef/ Gehlert, Günther/ Gülicher, Herbert, Statistik für Wirtschaftswissenschaftler, 15. Aufl., München 2008, S. 127.

¹²⁴ Vgl. Röbber, Irene/ Ungerer, Albrecht, Statistik für Wirtschaftswissenschaftler, 2. Aufl., Heidelberg, 2011, S. 165.

¹²⁵ Vgl. Bleymüller, Josef/ Gehlert, Günther/ Gülicher, Herbert, Statistik für Wirtschaftswissenschaftler, 15. Aufl., München, 2008, S. 101.

¹²⁶ Vgl. Eckstein, Peter, P., Repetitorium Statistik, 6. Aufl., Wiesbaden, 2006, S. 295ff.

¹²⁷ Vgl. Hochstädter, Dieter, Statistische Methodenlehre, 8. Aufl., Frankfurt a.M., 1996, S. 574.

Untersuchung gezeigt bzw. bewiesen werden soll. Wendet man den Chi²-Anpassungstest im Rahmen der Ziffernanalyse an, um die Übereinstimmung der empirischen Ziffernverteilung mit der unterstellten theoretischen Verteilung zu prüfen, so lautet diese „die Zufallsvariable X besitzt die Verteilung P_0 “ oder „ $H_0: = P_0$ “. Kommt es folglich zu einer Ablehnung von H_0 so kann ausgeschlossen werden, dass die empirischen Daten die angenommene Verteilung aufweisen.¹²⁸

Zur Ablehnung der Nullhypothese kommt es genau dann, wenn eine berechnete Testgröße innerhalb eines bestimmten Intervalls, dem sog. „Ablehnungsbereich“ K^* liegt, also $T \in K^*$ erfüllt ist. Letzterer stellt im Falle des Chi²-Anpassungstestes das halbseitig offene Intervall $(\chi^2_{f;1-\alpha}; \infty)$ dar.¹²⁹ Hierbei ist $\chi^2_{f;1-\alpha}$ ein Quantil der Chi²-Verteilung mit f Freiheitsgraden. Für f gilt $f = m - r - 1$, wobei m die Anzahl der Klassen und r die Anzahl geschätzter Parameter repräsentiert. α bezeichnet das Signifikanzniveau. In der Regel ist es im Falle der Ziffernanalyse nicht nötig, einzelne Parameter zu schätzen, sodass i.A. $r = 0$ angenommen werden kann. Die Verteilung von $\chi^2_{f;1-\alpha}$ ist in Anhang I dargestellt. Die Testgröße T wird im Falle des Chi²-Anpassungstestes mittels¹³⁰

$$T = \chi^2 = \sum_{j=0}^m \frac{(h_j - n \cdot p_j)^2}{n \cdot p_j} \quad (4)$$

berechnet. In dieser steht h_j für die Anzahl der Beobachtungswerte (empirische Häufigkeit) der Klasse j , n für den gesamten Stichprobenumfang und p_j für die hypothetische Klassenwahrscheinlichkeit der Klasse j , $n \cdot p_j$ bezeichnet auch die Erwartungshäufigkeit von j . Je größer folglich die Differenz zwischen der Erwartungshäufigkeit und der empirischen Häufigkeit ist, desto größer ist der Wert von T und je höher die Wahrscheinlichkeit, dass H_0 abgelehnt wird. Diese Differenz wird zur Berechnung von T sowohl quadriert als auch mittels der Erwartungshäufigkeit normiert. Das Quadrieren bewirkt hierbei zum einen die Verhinderung der Aufhebung positiver und negativer Abweichungen, zum anderen erfolgt eine höhere Gewichtung stärkerer Abweichungen. Letzteres wird durch die o.g. Normierung unterstützt: Besonders große Abweichungen finden bei kleinen Häufigkeiten stärkeres Gewicht.¹³¹ Ein Wert von $T = 0$

¹²⁸ Vgl. Polasek, Wolfgang, Schließende Statistik, Berlin, 1997, S. 105-109.

¹²⁹ Vgl. Bankhofer, Udo/ Vogel, Jürgen, Datenanalyse und Statistik, Wiesbaden, 2008, S. 131.

¹³⁰ Vgl. Alt, Raimund, Statistik – Eine Einführung für Wirtschaftswissenschaftler, 1. Aufl., Wien, 2010, S. 330.

¹³¹ Vgl. Trede, Marc/ Watrin, Christoph/ Ullmann, Robert, Ziffernanalyse und Chi-Quadrat-Anpassungstest in der steuerlichen Anwendung, in: Die Betriebswirtschaft, 2009, S. 705.

würde nur auftreten, wenn die beobachtete Verteilung exakt mit der theoretisch angenommenen Verteilung zusammenfällt.¹³²

Vorteilhaft an den nicht-parametrischen Testverfahren wie dem Chi²-Anpassungstest ist v.a. die Tatsache, dass eine Übereinstimmung der empirischen Verteilung mit jeder beliebigen theoretischen Verteilung überprüft werden kann.¹³³ Demzufolge eignen sich diese insbesondere auch, um die Benford-Verteilung im Rahmen der Ziffernanalyse zu Grunde zu legen.

Damit der Chi²-Anpassungstest jedoch Anwendung finden kann, müssen drei Voraussetzungen erfüllt sein.

Die erste Bedingung setzt voraus, dass die Daten nicht durch systematische Muster entstanden sein dürfen. Sie müssen zwingend zufälligen Ursprungs sein.¹³⁴ Erfüllt wäre dies z.B. bei den Passagierzahlen einer Airline, die auf den Ferienreiseverkehr spezialisiert ist. Problematisch hingegen wäre die Bedingung bei Linienfluggesellschaften, wenn diese vollständig von Pendlern an stets gleichen Tagen genutzt wird: In diesem Falle würden sich auch die Passagierzahlen in regelmäßigen Abständen wiederholen.

Die zweite zu erfüllende Bedingung steht in Zusammenhang mit dem Stichprobenumfang. Für diesen muss ein Mindestwert an Daten vorliegen, damit korrekte Schlussfolgerungen aus dem Test gezogen werden können. Aus diesem Grund ist für jede Merkmalsklasse ein Minimum der Erwartungshäufigkeit notwendig, welche im Falle des Chi²-Anpassungstests nach herrschender Meinung der Literatur mindestens einen Wert von 5 annehmen sollte.¹³⁵ Ist diese Voraussetzung nicht erfüllt, ist es theoretisch möglich, zwei oder mehr Merkmalsklassen zusammenzulegen, um den Mindestwert dann für die zusammengelegte Klasse zu erreichen.¹³⁶ Im Falle der Ziffernanalyse wäre dies jedoch weniger sinnvoll, da so keinerlei Rückschlüsse auf die Häufigkeiten einzelner Ziffern gezogen werden könnten. Problematisch können jedoch auch zu große Stichprobenumfänge sein: Wie in Kapitel 4 gezeigt wird, neigt die Testgröße des Chi²-Anpassungstestes dazu, im Falle sehr großer Stichprobenumfänge bereits bei kleinsten Abweichungen

¹³² Vgl. Watrin, Christoph/ Ullmann, Robert, Ziffernanalyse in der steuerlichen Betriebsprüfung: Voraussetzungen, Funktionsweise, Anwendungsmöglichkeiten, in: Die Wirtschaftsprüfung, 2009, S. 101f.

¹³³ Vgl. Sosna, Christian, Statistische Ziffernanalyse - Teil I -, in: Steuerliche Betriebsprüfung, 09/2004, S. 252.

¹³⁴ Vgl. Wähnert, Andreas, Anwendbarkeit, Aussagekraft und Grenzen stochastischer Manipulationstests, in: Steuerliche Betriebsprüfung, 11/2008, S. 314.

¹³⁵ Vgl. Bomsdorf, Eckart, Wahrscheinlichkeitsrechnung und statistische Inferenz, 8. Aufl., Köln, 2002, S. 127.

¹³⁶ Vgl. Watrin, Christoph/ Ullmann, Robert, Ziffernanalyse in der steuerlichen Betriebsprüfung: Voraussetzungen, Funktionsweise, Anwendungsmöglichkeiten, in: Die Wirtschaftsprüfung, 2009, S. 102.

sehr große Werte anzunehmen und somit eine Nullhypothese statistisch signifikant abzulehnen. *Nigrini* hält jedoch Stichprobenumfänge bis $n = 10.000$ Daten für vollkommen unbedenklich.¹³⁷

Gemäß der dritten Voraussetzung muss die Stichprobe stets „einfach“ sein. Dies ist dann erfüllt, wenn alle Daten mit der gleichen Wahrscheinlichkeit in die Stichprobe gelangen und die Ziehung aus der Grundgesamtheit voneinander unabhängig erfolgt.¹³⁸ Problematisch ist hierbei immer, wenn einzelne Daten mittels mathematischer Operationen durch andere Daten berechnet werden können. Werden beispielsweise makroökonomische Daten untersucht, so lässt sich das Bruttoinlandsprodukt (BIP) u.a. aus den Größen der Bruttoinvestitionen und der Bruttowertschöpfung berechnen. Demzufolge dürften Daten des BIP nicht herangezogen werden, wenn ebenfalls Daten der Bruttowertschöpfung und der Bruttoinvestitionen vorliegen.

Nach Festlegung des Signifikanzniveaus α werden bei Durchführung des Chi²-Anpassungstestes äquivalent zu sämtlichen anderen nicht-parametrischen Testverfahren vier Schritte durchlaufen.¹³⁹

Zunächst einmal ist es notwendig, eine Zählung der empirischen Daten vorzunehmen und diese in Merkmalsklassen einzuordnen. Im Falle der Ziffernanalyse werden hierzu die Anfangsziffern der vorliegenden Zahlen extrahiert und den Merkmalsklassen in Form der Ziffern Eins bzw. Null bis Neun zugeordnet. Im Anschluss werden die Elemente in den Klassen gezählt und somit deren empirischen Häufigkeiten bestimmt.

Als zweites werden die Erwartungshäufigkeiten der Merkmalsklassen in Abhängigkeit der unterstellten Verteilungsfunktion und des Stichprobenumfanges bestimmt. Bei Anwendung in der Ziffernanalyse sind dies die Häufigkeiten, die bei nicht-manipulierten Daten zu erwarten wären. Sind alle Erwartungshäufigkeiten bestimmt, kann unmittelbar geprüft werden, ob der o.g. Mindestwert für alle Klassen eingehalten wird.

¹³⁷ Vgl. Nigrini, Mark, *Digital Analysis Using Benford's Law: Test & Statistics for Auditors*, 2. Aufl., Vancouver, 2000, S. 78.

¹³⁸ Vgl. Büning, Herbert/ Trenkler, Götz, *Nichtparametrische statistische Methoden*, 2. Aufl., Berlin et al., 1998, S. 75.

¹³⁹ Vgl. Watrin, Christoph/Ullmann, Robert, *Ziffernanalyse in der steuerlichen Betriebsprüfung: Voraussetzungen, Funktionsweise, Anwendungsmöglichkeiten*, in: *Die Wirtschaftsprüfung*, 02/2009, S. 102.

In einem dritten Schritt wird aus den empirischen Häufigkeiten und den Erwartungshäufigkeiten die Testgröße T bestimmt, welche zeitgleich das Maß für die Abweichung der empirischen Daten von der unterstellten Verteilung darstellt.

Zum Schluss gilt es noch den Ablehnungsbereich K^* zu bestimmen. Hierzu werden zunächst die Freiheitsgrade aus der Anzahl der Merkmalsklassen und der Zahl geschätzter Parameter ermittelt und im Anschluss der kritische Wert aus der χ^2 -Verteilung abgelesen. Somit kann geprüft werden, ob die Testgröße innerhalb des Ablehnungsbereiches liegt oder nicht und somit entschieden werden, ob die Nullhypothese abzulehnen ist.

Um die Vorgehensweise des χ^2 -Anpassungstestes praktisch zu erläutern, könnte man mittels diesem einen Roulette-Tisch eines Kasinos auf Manipulationen prüfen. Dieser besteht aus den Zahlen 1 bis 36, die jeweils abwechselnd in den Farben rot und schwarz vorkommen und einer grünen Null. Demnach können 37 verschiedene Ereignisse eintreten. Die Spieler können dann auf den Eintritt eines bestimmten Ereignisses, z.B. das Auftreten der Farbe oder einer bestimmten Zahl bzw. eines Zahlenbereiches (Zahl 1 bis 18 und 19 bis 36) setzen. Tritt die grüne Null auf, so gewinnt die Spielbank, also das Kasino. Die Wahrscheinlichkeit für das Auftreten einer bestimmten Zahl ist hierbei stets gleich und beträgt $\frac{1}{37} \approx 2,7 \%$.

Möchte ein Spieler nun wissen, ob die Spielbank das Roulette manipuliert hat, liegt es nahe zu überprüfen, ob diese zu häufig gewinnt, also ob die Zahl Null zu häufig auftritt. Demzufolge lautet die Nullhypothese zu dieser Fragestellung, „Die auftretenden Zahlen sind gleichverteilt mit der Wahrscheinlichkeit von 2,7 %“. Beträgt das Signifikanzniveau $\alpha = 0,95$, ergibt sich der Ablehnungsbereich zu $K^* = (51,0; \infty)$.

Spielt man nun 185 Runden, so müsste idealerweise jede Zahl exakt fünf Mal auftreten. Tritt die Zahl Null zehn Mal auf, so ergibt sich ein χ^2 -Wert der Testgröße von 6,0. Die Nullhypothese wäre nicht abzulehnen. Gewinnt das Kasino 41 Mal, so würde der χ^2 -Wert bei 266,4 liegen, was eindeutig in dem Ablehnungsbereich K^* liegt, sodass H_0 signifikant verworfen würde und die Gleichverteilung abgelehnt werden müsste. Die Wahrscheinlichkeit einer Manipulation wäre in diesem Fall augenscheinlich sehr groß. Tritt die Null in 185 Würfeln exakt 21 Mal auf, so ist dies die geringste Anzahl an Kasinogewinnen, bei der es erstmals zu einer Ablehnung der Nullhypothese kommt. Der χ^2 -Wert beträgt 54,4. Hierbei kann mit einer Wahrscheinlichkeit von 95% davon ausgegangen werden, dass Manipulationen vorhanden sind.

Senkt man das Signifikanzniveau auf $\alpha = 0,01$ würde die Nullhypothese erst ab einem kritischen Chi²-Wert von 58,62 abgelehnt werden. Hierzu wären dann 22 Einsen notwendig, wobei die Irrtumswahrscheinlichkeit für eine falsche Ablehnung lediglich 1% beträgt. Spielt man hingegen 1.850 Mal, so beträgt der Erwartungswert für die Zahl Null 50. Bei einem Signifikanzniveau von $\alpha = 0,05$ sind 100 Nullen für eine signifikante Ablehnung der Gleichverteilung notwendig, welche einen Chi²-Wert von 51,56 implizieren, der den kritischen Wert von 51,00 nur knapp überschreitet.

Wie bereits dargelegt, können jedoch auch im Falle einer signifikanten Entscheidung Fehler der ersten und zweiten Art auftreten. Dies ist gerade im Falle der Manipulationsaufdeckung problematisch, da so u.U. falsche Manipulationsvorwürfe erhoben werden. Aus diesem Grund sollte bei allen statistischen Testverfahren zwingend zwischen statistischer und ökonomischer Signifikanz unterschieden werden.¹⁴⁰ Eine weitere Vorgehensweise, welche sich deshalb in der Vergangenheit in einer Reihe von Forschungsarbeiten durchgesetzt hat, ist der primäre Fokus auf die ökonomische Signifikanz: Die Ergebnisse werden in diesem Fall relativ betrachtet, was die Bildung von Rangfolgen hinsichtlich der Höhe des Chi²-Wertes der zu untersuchenden Objekte zur Folge hat. Hinsichtlich der Platzierung innerhalb des Rankings können dann Rückschlüsse bzgl. der Wahrscheinlichkeit vorhandener Manipulationen getroffen werden.

3.4.2 Kolmogorow-Smirnow-Anpassungstest

Ein weiteres nicht-parametrisches Testverfahren ist der Kolmogorow-Smirnow-Anpassungstest (KS-Test). Die grundlegende Funktionsweise ist hierbei ähnlich zu der des Chi²-Anpassungstestes: Ziel ist es, eine im Vorfeld formulierte Nullhypothese abzulehnen, wobei diese äquivalent zu derjenigen des Chi²-Anpassungstestes ist und somit ebenso „Der Datensatz folgt der Verteilung P_0 “ lautet.¹⁴¹ Diese wird auch hier abgelehnt, sofern die errechnete Testgröße innerhalb des ermittelten Ablehnungsbereiches liegt. Die Testgröße wird auf Basis des maximalen Abstandes zwischen der empirischen Verteilung und der unterstellten Verteilung berechnet. Sie lautet:

$$T = D = \max_x |F_0(x) - F_e(x)|. \quad (5)$$

¹⁴⁰ Vgl. Krämer, Walter, The Cult of Statistical Significance – What Economists Should and Should Not Do to Make their Data Talk, in: Schmollers Jahrbuch, 2011, S. 455 - 468.

¹⁴¹ Vgl. Hafner, Robert, Nichtparametrische Verfahren der Statistik, Wien, 2001, S. 182.

Hierbei bezeichnet $F_0(x)$ die unterstellte Verteilung und $F_e(x)$ die Verteilungsfunktion der empirischen Beobachtungswerte. Hierbei wurde der KS-Test im eigentlichen Sinne für stetige Verteilungsfunktionen konzipiert.¹⁴² Auch wenn der Einsatz für diskrete Verteilungen grundsätzlich möglich ist, nimmt hierbei die Güte des Testverfahrens ab, da die Trennschärfe des Verfahrens abnimmt und somit die Nullhypothese in weniger Fällen abgelehnt wird als es bei stetigen Verteilungen der Fall ist.¹⁴³

Wird der KS-Test in der Praxis angewendet, wird zunächst unterstellt, dass die n Beobachtungswerte x_i ($i = 1, \dots, n$) in aufsteigender Form ($x_1 \leq x_2 \leq \dots \leq x_n$) vorliegen. Mittels dieser wird dann die empirische Verteilungsfunktion als relative Summenfunktion $S(x_i) = F_e(x_i)$ gebildet. Gleichzeitig wird die hypothetische Verteilung $F_0(x_i)$ jeweils an den Stellen x_i ausgewertet und der Abstand zu der empirischen Verteilungsfunktion ermittelt. Der insgesamt größte auftretende Abstand entspricht dann der Testgröße T .

Der kritische Wert, welcher für eine Ablehnung der Nullhypothese überschritten werden muss, ist bis zu einem Stichprobenumfang von $n = 35$ tabelliert.¹⁴⁴ Im Falle größerer Stichprobenumfänge wird dieser mittels einer Näherungsformel berechnet, die abhängig vom Signifikanzniveau ist. Eine Übersicht für stetige Verteilungen sowie eine von *Morrow* entwickelte Näherung bei Zugrundelegung der Benford-Verteilung gibt hierbei Tabelle 2.

	$\alpha = 0,1$	$\alpha = 0,05$	$\alpha = 0,01$
Stetige Verteilung	$1,224/\sqrt{n}$	$1,358/\sqrt{n}$	$1,628/\sqrt{n}$
Benford-Verteilung	$1,012/\sqrt{n}$	$1,148/\sqrt{n}$	$1,420/\sqrt{n}$

Tabelle 2: Näherungsformeln für den kritischen Wert des KS-Tests bei stetigen Verteilungen sowie der Benford-Verteilung¹⁴⁵

Im Vergleich zum Chi²-Anpassungstest eignet sich der KS-Test auch für kleinere Stichprobengrößen. Grundsätzlich können hierbei sämtliche stetige und diskrete Verteilungen unterstellt

¹⁴² Vgl. Bleymüller, Joseph/ Gehlert, Günther/ Gülicher, Herbert, Statistik für Wirtschaftswissenschaftler, 15. Auflage, München, 2008, S. 127.

¹⁴³ Vgl. Bünning, Herbert/ Trenkler, Götz, Nichtparametrische statistische Methoden, 2. Auflage, Berlin, 1994, S. 72.

¹⁴⁴ Vgl. Distance Learning Center, Appendix – Table 7: Kolmogorov-Smirnov Test, <http://dlc.eriei.com/onlinetextbook/index.cfm?fuseaction=textbook.appendix&FileName=Table7> (Abgerufen am 16.03.2015), 2014.

¹⁴⁵ Vgl. Morrow, John, Benford's Law, Families of Distributions and a Test Basis, <http://johnmorrow.info/pdf/benford.pdf> (Abgerufen am 17.12.2014), 2014, S. 4.

werden und somit auch die Benford-Verteilung, wodurch der Test vielseitige Anwendungsfelder bietet. So setzten auch *Nigrini* und *Mittermaier* den Test als Ergänzung zum χ^2 -Anpassungstest ein.¹⁴⁶

Wendet man den KS-Test in dem unter 3.4.1 angeführten Beispiel eines Roulette-Tisches an, so lautet auch hierbei die Nullhypothese „Die auftretenden Zahlen folgen einer Gleichverteilung“. Abhängig vom Stichprobenumfang ergeben sich gemäß Tabelle 2 für ein Signifikanzniveau von 0,05 unterschiedliche kritische Werte. Für $n = 185$ ergibt sich der kritische Wert zu $\frac{1,358}{\sqrt{185}} = 0,0998$. Betrachtet man erneut die Anzahl an Nullen, so sollte deren Anzahl idealerweise fünf betragen. Erstmalig übersteigt die Testgröße mit einem Wert von 0,103 im Falle von 24 Nullen den kritischen Wert, sodass die Nullhypothese mit der einer Wahrscheinlichkeit von 95% korrekt abgelehnt wird und vorhandene Manipulationen angenommen werden können. Im Falle von $n = 1.850$ beträgt der kritische Wert 0,0315, idealerweise müsste die Anzahl an Nullen 50 betragen. Abzulehnen ist die Nullhypothese dann im Falle von 109 Kasinogewinnen, wobei $T = 0,0318$ gilt.

3.4.3 Kuiper-Test

Auch der Kuiper-Test stellt ein nicht-parametrisches Testverfahren dar, sodass die Vorgehensweise derjenigen des χ^2 -Anpassungstests bzw. des KS-Tests entspricht: Die Nullhypothese $H_0 := P_0$ kann nur statistisch signifikant abgelehnt werden. Hierzu muss eine Testgröße innerhalb eines Ablehnungsbereiches K^* liegen. Prinzipiell basiert der Test auf dem gleichen Prinzip wie der KS-Test: Auch hier berechnet sich die Testgröße auf Basis der maximalen Differenz zwischen empirischer und unterstellter Verteilungsfunktion. Im Unterschied wird hierbei jedoch nicht der Betrag herangezogen, sondern die maximale positive und die maximale negative Differenz aufaddiert. Demnach lautet die Testgröße

$$T = V_N = \max_x (F_0(x) - F_e(x)) + \max_x (F_e(x) - F_0(x)).^{147} \quad (6)$$

¹⁴⁶ Vgl. Nigrini, Mark/ Miller, Steven, Benford's Law applied to Hydrological Data-Results and Relevance to other Geophysical Data, in: Journal of Mathematical Geology, 2007, Nr. 5, S. 477.

¹⁴⁷ Vgl. Giles, David E., Benford's Law and naturally occurring prices in certain eBay auctions, in: Applied Economic Letters, 2007, Vol. 14, S. 160.

Stephens modifizierte diesen Ansatz, sodass das Testverfahren unabhängig vom Stichprobenumfang ist. Diese modifizierte Testgröße berechnet sich gemäß

$$T^* = V_N \cdot \left[\sqrt{n} + 0,155 + \frac{0,24}{\sqrt{n}} \right].^{148} \quad (7)$$

Auch lieferte *Stephens* hierbei für den modifizierten Test im Falle stetiger Verteilungen kritische Werte für verschiedene Signifikanzniveaus. *Morrow* publizierte speziell auf die Benford-Verteilung zugeschnittene Werte. Diese Anpassung ist wie auch beim KS-Test nötig, da der Test grundsätzlich für stetige Verteilungen konzipiert wurde und im Falle diskreter hypothetischer Verteilungen andernfalls falsche Testergebnisse liefern würde. Die kritischen Werte für stetige Verteilungen und die Benford-Verteilung zeigt Tabelle 3.

	$\alpha = 0,1$	$\alpha = 0,05$	$\alpha = 0,01$
Stetige Verteilung	1,620	1,747	2,001
Benford-Verteilung	1,191	1,321	1,579

Tabelle 3: Näherungsformeln für den kritischen Wert des Kuiper-Tests bei stetigen Verteilungen sowie der Benford-Verteilung¹⁴⁹

Aufgrund der Invarianz bzgl. des Stichprobenumfanges eignet sich der Test der Theorie nach auch für sehr große Stichprobenumfänge.¹⁵⁰

Wendet man auch diesen Test zur Manipulationsaufdeckung im Falle des Roulette-Tisches an, so wäre die Nullhypothese im Falle von 185 Würfeln selbst im Falle von $\alpha = 0,1$ nicht abzulehnen, wenn die Anzahl an Nullen 24 beträgt. Hier läge der Wert von $T = V_N^*$ bei 1,415 und somit unter dem kritischen Wert von 1,620. Bei einem Signifikanzniveau von 0,95 wären insgesamt 29 Bankgewinne notwendig, um den kritischen Wert von 1,747 zu überschreiten, wobei die Testgröße bei einem Wert von 1,787 liegt. Hier käme man mit einer Irrtumswahrscheinlichkeit von 5% zu dem Schluss, dass das Roulette manipuliert worden ist. Bei 32 Nullen würde mit einer Testgröße von 2,010 sogar der kritische Wert für $\alpha = 0,01$ überschritten, sodass sich die Irrtumswahrscheinlichkeit auf 1% reduziert.

¹⁴⁸ Vgl. Stephens, M. A., The goodness-of-fit statistic, VN: Distribution and significance points, in: Biometrika, 1965, Vol. 52, S. 309–321.

¹⁴⁹ Vgl. Stephens, M. A., The goodness-of-fit statistic, VN: Distribution and significance points, in: Biometrika, 1965, Vol. 52, S. 309–321 und Morrow, John, Benford's Law, Families of Distributions and a Test Basis, <http://johnmorrow.info/pdf/benford.pdf> (Abgerufen am 17.12.2014), 2014, S. 4.

¹⁵⁰ Vgl. Giles, David E., Benford's Law and naturally occurring prices in certain eBay auctions, in: Applied Economic Letters, 2007, Vol. 14, S. 160.

3.4.4 Test der Z-Statistik

Im Gegensatz zu den bisher vorgestellten Verfahren trifft die Z-Statistik die Aussagen hinsichtlich der Konformität empirischer Daten mit einer hypothetischen Verteilung nicht auf Basis der Einbeziehung sämtlicher vorliegender Klassen. Stattdessen wird für jede Klasse j die Abweichung der relativen empirischen Häufigkeit f_j von der hypothetischen Klassenwahrscheinlichkeit p_j berechnet und in eine näherungsweise normalverteilte Zufallszahl Z umgewandelt. Diese berechnet sich mittels

$$Z = \frac{|f_j - p_j| - \left(\frac{1}{2n}\right)}{\sqrt{\frac{p_j(1-p_j)}{n}}} \quad (8)$$

Der Term $\left(\frac{1}{2n}\right)$ im Zähler der o.g. Formel stellt hierbei eine Stetigkeitskorrektur dar. Diese ist notwendig, wenn die zu analysierenden Daten eine diskrete Verteilung aufweisen bzw. die unterstellte Verteilung diskret ist. Da die Normalverteilung stetig ist, muss eine entsprechende Korrektur vorgenommen werden. Ist der entsprechende Wert von Z berechnet, so gilt es im Anschluss zu prüfen, ob dieser größer als der kritische Wert ist und damit die im Vorfeld formulierte Nullhypothese „Der Datensatz folgt der Verteilung P_0 “ abgelehnt werden kann oder nicht. Eine Übersicht der kritischen Werte mit unterschiedlichen Signifikanzniveaus zeigt Tabelle 4.

	$\alpha = 0,1$	$\alpha = 0,05$	$\alpha = 0,01$
Kritische Werte	1,645	1,960	2,576

Tabelle 4: Kritische Werte der Z-Statistik¹⁵²

Auch bei diesem Test kann grundsätzlich jede Verteilung P_0 unterstellt werden, sodass das Verfahren auch im Rahmen des Benford-Testes Anwendung finden kann. Hierbei entspricht jede vorkommende Ziffer einer Klasse. Demzufolge kann dann für jede Ziffer isoliert die Aussage getroffen werden, ob für diese die Benford-Verteilung abzulehnen ist oder nicht. Fraglich ist in diesem Zusammenhang, ob im Falle der Ablehnung der hypothetischen Verteilung einer Ziffer, diese für den gesamten Datensatz abgelehnt werden kann oder nicht. *Nigrini* beobachtete in diesem Zusammenhang, dass dies nicht unbedingt der Fall sein muss: Erst wenn mehrere

¹⁵¹ Vgl. Nigrini, Mark, Digital Analysis Using Benford's Law, Vancouver, 2000, S. 42.

¹⁵² Vgl. Bankhofer, Udo/ Vogel, Jürgen, Datenanalyse und Statistik, Wiesbaden, 2008, S. 101.

Ziffern signifikante Abweichungen aufweisen, sei auch im Falle anderer Verfahren eine Ablehnung der Verteilung zu beobachten.¹⁵³

Augenscheinlich ist für eine solche Aussage jedoch auch zu berücksichtigen, wie stark die Abweichungen der einzelnen Klassen sind. Demnach ist es theoretisch sehr wohl möglich, dass bei großen Deviationen einer einzelnen Klasse auch andere Testverfahren die hypothetische Verteilung ablehnen. Veranschaulicht werden soll dies im Falle des Beispiels des Kasino-Roulettes. Aus dem χ^2 -Anpassungstest beispielsweise resultierte eine Ablehnung der Nullhypothese „Die vom Roulette hervorgebrachten Zufallszahlen sind gleichverteilt“ mit $\alpha = 0,05$ im Falle des Auftretens von 21 Nullen bei 185 Durchgängen. Hierbei liegt bereits nahe, dass die Z-Statistik eine Ablehnung der Nullhypothese für die Zahl Null hervorbringen wird. Für die Klasse der Zahl Null ergibt sich ein Wert von $Z = 7,027$, der deutlich über dem kritischen Wert von 1,96 bei $\alpha = 0,05$ liegt. Für alle anderen Klassen ergibt sich keine Überschreitung des kritischen Wertes.

Dieses Beispiel zeigt einen weiteren Vorteil der Z-Statistik: Mit dieser ist es im Falle einer mittels eines anderen Verfahrens gezeigten Abweichung möglich, die hierfür verantwortliche Klasse zu identifizieren. Bei Anwendung in der Ziffernanalyse lässt sich somit genau dokumentieren, welche Ziffern signifikante Deviationen aufweisen und welche konform mit der hypothetischen Verteilung sind.

3.4.5 Normierter χ^2 -Wert

Wie oben bereits erläutert, besitzt der χ^2 -Anpassungstest die negative Eigenschaft, dass die Testgröße $T = \chi^2$ bei steigendem Stichprobenumfang und geringen Deviationen ebenfalls stark ansteigt. Dies sorgt zum einen dafür, dass bei großen Stichprobenumfängen häufig falsch positive Testergebnisse auftreten. Zum anderen erschwert dies die Vergleichbarkeit zweier Stichproben, sofern sich deren Stichprobenumfänge unterscheiden. Aus diesem Grund nutzen einige Forscher die Testgröße χ^2 dividiert durch den Stichprobenumfang, um diese negative Eigenschaft auszugleichen und die Vergleichbarkeit der Ergebnisse zu gewährleisten.¹⁵⁴ Für diese Kenngröße gilt also:

¹⁵³ Vgl. Nigrini, Mark, Digital Analysis Using Benford's Law, Vancouver, 2000, S. 43.

¹⁵⁴ Vgl. Leemis, Lawrence M./ Schmeiser, Bruce W./ Evans, Diane L., Survival Distributions Satisfying Benford's Law, in: American Statistician, 2000, S. 236-241.

$$c = \frac{\chi^2}{n} \cdot 155 \quad (9)$$

Hierbei gilt $0 \leq c \leq 1$. Nachteilig bei Nutzung dieser Größe ist allerdings, dass prinzipiell keine kritischen Werte existieren, ab welchen eine signifikante Abweichung von der unterstellten Verteilung anzunehmen ist. Folglich ist es lediglich möglich, zwei Stichproben hinsichtlich ihres Abweichungsgrades zu vergleichen.

Eine beispielhafte Anwendung wäre z.B. der Vergleich zweier Roulettetische in einem Kasino hinsichtlich ihrer Manipulationswahrscheinlichkeit. Hierzu würde man beispielsweise 185 Runden spielen und die Zahlen notieren. Treten dann bei Roulette A 21 Nullen auf, so beträgt der Wert $c = 0,2941$. Bringt Roulette B nur 18 Nullen hervor, ergibt dies einen Wert von $c = 0,1968$. Da letzterer geringer ausfällt als der des Roulettes A, wäre die Wahrscheinlichkeit vorhandener Manipulationen hier deutlich geringer. Demzufolge sollte Roulette A einer genaueren Untersuchung unterzogen werden.

3.4.6 Normierter euklidischer Abstand

Die Kenngröße des normierten euklidischen Abstandes gehört zu der Gruppe der sog. „Abstandsmaße“. Mathematisch wird bei diesen der Abstand der empirischen Verteilung von der angenommenen theoretischen Verteilung berechnet. Je größer dieser ausfällt, desto geringer ist die Wahrscheinlichkeit, dass die empirischen Daten tatsächlich der unterstellten Verteilung folgen.

Berechnet man den euklidischen Abstand zwischen zwei Punkten x_i und x_j in der Ebene, so bezeichnet dieser den kürzesten Abstand der Punkte, welchen man ebenfalls mit dem Lineal nachmessen könnte. Verbildlicht ist dies in Abbildung 3 anhand des Satzes von Pythagoras: Die Länge der Hypotenuse c wird auf Basis der zwei Katheten eines rechtwinkligen Dreiecks mittels

$$c = \sqrt{a^2 + b^2} = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2} \quad (10)$$

¹⁵⁵ Hieraus lässt sich die Invarianz bzgl. des Stichprobenumfanges begründen, denn es gilt: $c = \frac{\chi^2}{n} = \frac{1}{n} \cdot \sum_{j=0}^m \frac{(h_j - n \cdot p_j)^2}{n \cdot p_j} = \frac{n^2}{n^2} \cdot \sum_{j=0}^m \frac{(f_j - p_j)^2}{p_j} = \sum_{j=0}^m \frac{(f_j - p_j)^2}{p_j}$.

berechnet.¹⁵⁶

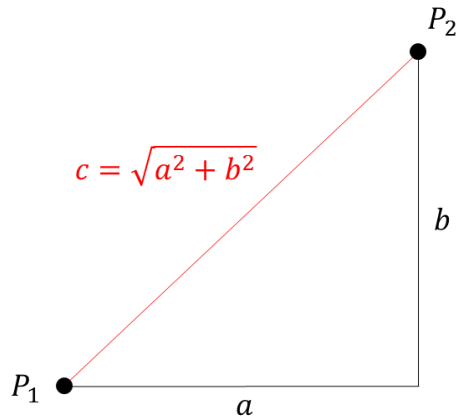


Abbildung 3: Berechnung des Abstandes zweier Punkte mittels des Satzes des Pythagoras

Erweitert man das Modell auf N Dimensionen, so ergibt sich der euklidische Abstand zu:¹⁵⁷

$$d_{ij} = \|x_i - x_j\| = \sqrt{\sum_{n=1}^N (x_{in} - x_{jn})^2} \quad (11)$$

Selbiges Prinzip findet auch in der verallgemeinerten Berechnung des euklidischen Abstandes im Rahmen der statistischen Verwendung Anwendung. Hierbei wird entsprechend der Abstand zwischen den Klassenmitten der empirischen sowie der hypothetischen Verteilung quadriert und aufsummiert. Im Anschluss wird die Wurzel aus dem berechneten Wert gezogen. Mittels einer Gleichung ausgedrückt, ergibt sich der euklidische Abstand somit zu

$$d = \sqrt{\sum_{j=1}^m (f_j - p_j)^2}. \quad (12)$$

Hierbei bezeichnet f_j die relative Häufigkeit und p_j die hypothetische Wahrscheinlichkeit der Klasse j . Dividiert man diesen Abstand durch die maximal mögliche Distanz, so hat dies den Vorteil, dass damit auch Abstände zu unterschiedlichen Verteilungen verglichen werden können.¹⁵⁸ Aus diesem Grund erscheint es als sinnvoll, eine entsprechende Normierung vorzunehmen, sodass sich die Berechnung des normierten euklidischen Abstandes zu

¹⁵⁶ Vgl. Eckstein, Peter, Statistik für Wirtschaftswissenschaftler, Wiesbaden 2008, S. 379f.

¹⁵⁷ Vgl. Schlittgen, Rainer, Multivariate Statistik, München 2009, S. 295-297, 532.

¹⁵⁸ Vgl. Cho, Tam/ Wendy K./ Gaines, Brian J., Breaking the (Benford) Law: Staistical Fraud Detection in Campaign Finance, in: American Statistician, 2007, S. 221.

$$d^* = \sqrt{\frac{\sum_{j=1}^m (f_j - p_j)^2}{\sum_{j=1}^{m-1} p_j^2 + (1 - p_m)^2}} \quad (13)$$

ergibt. Es gilt hierbei immer $d^* \in [0,1]$.

Wie bereits im Falle des normierten Chi²-Wertes ist es jedoch mit dieser Testgröße nicht möglich, pauschale Aussagen darüber zu treffen, ob tatsächlich eine Abweichung von der theoretischen Verteilung vorliegt oder nicht. Demnach können Ergebnisse auch hier lediglich relativ betrachtet werden.

Auf das Beispiel des Vergleiches zweier Roulettetische zurückkommend, würde sich hierbei im Falle von 21 Nullen des Roulettes A ein Wert von $d^* = 0,1563$ ergeben. Werden mit Tisch B erneut lediglich 18 Nullen generiert, resultiert daraus der Wert $d^* = 0,0739$, sodass man erneut zu dem Schluss käme, Roulette A ist mit einer höheren Wahrscheinlichkeit Manipulationen unterworfen worden.

3.4.7 Abstand über die arithmetischen Mittelwerte

Ein weiteres Abstandsmaß ist der sog. „Abstand über die arithmetischen Mittelwerte“. Bei diesem werden zunächst die arithmetischen Mittelwerte der hypothetischen sowie der empirischen Verteilung berechnet und voneinander abgezogen. Der Betrag dieser Differenz wird im Anschluss durch den größtmöglichen Abstand dividiert, welcher aus dem Saldo zwischen der Klassenanzahl und dem arithmetischen Mittel der hypothetischen Verteilung berechnet wird. Auch dies dient im Wesentlichen der Normierung der berechneten Größe, sodass eine bessere Vergleichbarkeit der Ergebnisse gegeben ist. Formal lässt sich die Kennzahl mittels

$$a^* = \frac{|\mu_e - \mu_b|}{m - \mu_b} \quad (14)$$

berechnen. Hierbei bezeichnet μ_e den Mittelwert der empirischen Daten, μ_b den Mittelwert der hypothetischen Verteilung und m die Anzahl der Klassen. Im Falle der Benford-Verteilung beträgt das arithmetische Mittel der Verteilung für die ersten Ziffern

$$\mu_b = 1 \cdot 0,301 + 2 \cdot 0,176 + 3 \cdot 0,125 + 4 \cdot 0,097 + \dots + 9 \cdot 0,046 = 3,4407.$$

Für die zweiten Ziffern beträgt der zu erwartende Mittelwert

$$\mu_b = 0 \cdot 0,1196 + 1 \cdot 0,1139 + 2 \cdot 0,1088 + \dots + 9 \cdot 0,0850 = 4,1874.$$

Auch für a^* gilt $0 \leq a^* \leq 1$. Hierbei ist der größtmögliche Abstand $a^* = 1$ gegeben, wenn alle Ziffern der empirischen Daten mit einer neun beginnen.¹⁵⁹

Bei ebenfalls 185 Roulette-Runden an zwei Tischen und 21 Nullen im Falle des ersten Roulettes würde sich ein Wert von $a^* = 0,0387$ ergeben. Bringt Roulette B erneut 18 Nullen hervor, so liegt auch hier der Wert $a^* = 0,0370$ unterhalb des Wertes des Roulettes A. Demnach wäre auch hier die Manipulationswahrscheinlichkeit im Falle des ersten Roulettes höher.

3.4.8 Manhattan-Distanz und Mean-Absolute-Deviation

Die Manhattan-Distanz wird auch als sog. „City-Block-Distanz“ oder „Taxi-Distanz“ bezeichnet. Berechnet wird diese auf Basis der Aufsummierung der Beträge der Abweichungen einzelner Koordinaten bzw. Klassenmitten. Formeltechnisch lässt sich diese mittels

$$MD = \sum_{j=1}^m |f_j - p_j| \quad (15)$$

beschreiben, wobei f_j die relativen empirischen und p_j die hypothetischen Häufigkeiten von Klasse j darstellen. Die Namensgebung dieses Distanzmaßes resultiert aus der schachbrett-artigen Anordnung der Wohnblöcke innerhalb des New Yorker Stadtteils Manhattan und dem auf Basis dessen orthogonal angeordneten Straßennetzes. Hierbei können z.B. Taxifahrer, wenn sie von A nach B wollen, die Wohnblöcke nicht diagonal kreuzen, sondern sind darauf angewiesen, die vertikalen und horizontalen Straßenstücke zu nutzen. Illustriert wird dies in Abbildung 4.

¹⁵⁹ Vgl. Judge, George/ Schechter, Laura, Detecting Problems in Survey Data Using Benford's Law, in: Journal of Human Resources, 2009, S. 6.

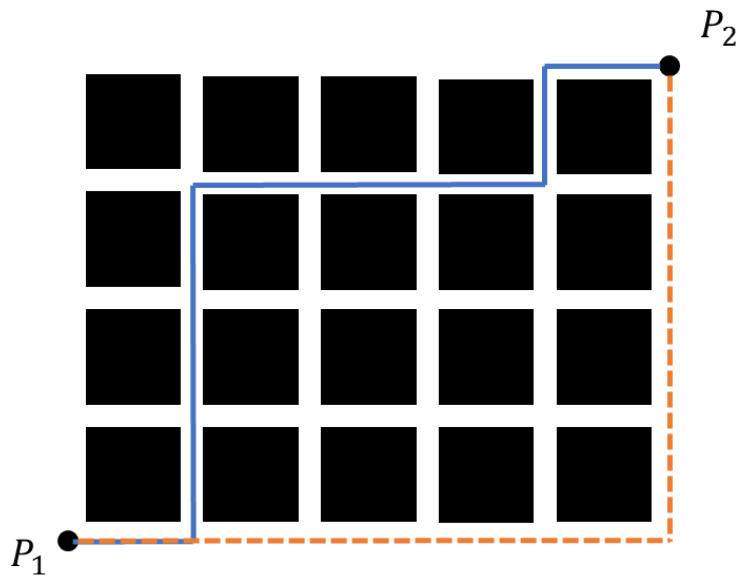


Abbildung 4: Grafische Illustration der Manhattan-Distanz zwischen zweier Punkte

Hierbei ist die Distanz stets unabhängig von dem gewählten „Weg“ durch die Straßen. Bei statistischer Verwendung liegt grundsätzlich das gleiche Prinzip zu Grunde: Hier werden die relativen Abweichungen zwischen den empirischen und den hypothetischen Klassenwahrscheinlichkeiten bestimmt und aufsummiert.¹⁶⁰

Findet auch hier das Beispiel mit dem Vergleich zweier Roulette-Tische Anwendung, so erhält man bei den 21 Nullen des ersten Tisches den Wert $MD = 0,1730$. Bei 18 Nullen des Roulettes B hingegen erhält man einen Wert von $MD = 0,1405$, sodass auch hier von einer geringeren Manipulationswahrscheinlichkeit des zweiten Würfels ausgegangen werden muss.

Anlehnend an diese Kennzahl existiert des Weiteren die sog. „Mean-Absolute-Deviation“. Diese gibt die mittlere absolute Abweichung der Klassen an und wird u.a. von *Nigrini* zur Ermittlung der Abweichung von der Benford-Verteilung verwendet.¹⁶¹ Berechnen lässt sich diese entsprechend gemäß

$$MAD = \frac{\sum_{j=1}^m |f_j - p_j|}{m}. \quad (16)$$

Nigrini lieferte hierbei Richtwerte zur Beurteilung der Abweichung von der hypothetischen Verteilung, welche in Tabelle 5 dargestellt sind.

¹⁶⁰ Vgl. Schlittgen, Rainer, Multivariate Statistik, München, 2009, S. 64.

¹⁶¹ Vgl. Nigrini Mark, Benford's Law, New Jersey 2012, S. 158f.

Interpretation	<i>MAD</i>
nahezu übereinstimmend	[0; 0,004)
noch übereinstimmend	[0,004; 0,008)
geringfügig übereinstimmend	[0,008; 0,012)
nicht übereinstimmend	[0,012, ∞)

Tabelle 5: Richtwerte zur Beurteilung der *MAD*¹⁶²

Im Falle der Roulette-Tische müssten die Ergebnisse der Manhattan-Distanz noch durch die Anzahl der Merkmalsklassen dividiert werden, um den Wert der *MAD* zu erhalten. Da sich die Zahlen 0 bis 36 ergeben können, beträgt diese 37. Demnach ergibt sich für Tisch A ein Wert von $MAD = 0,00468$ und für Tisch B $MAD = 0,00380$, was gemäß den Richtwerten in beiden Fällen eine Abweichung von der Gleichverteilung signalisieren würde.

Da sich die Größen *MAD* und *MD* nur durch die Division der Merkmalsklassen unterscheiden, ist es ausreichend, lediglich eine der beiden Größen in den Untersuchungen zu berücksichtigen, solange die Zahl der Merkmalsklassen stets gleich ist. Da dies bei dem Benford-Test gegeben ist, wird im Folgenden lediglich die Größe *MD* näher untersucht.

3.4.9 Tschebyscheff-Distanz

Die Tschebyscheff-Distanz gibt den maximalen Abstand zweier Objekte entlang einer Achse im m-dimensionalen Zahlenraum an. Sie berechnet sich mittels

$$TD = \max_j |f_j - p_j|. \quad (17)$$

Wird dieses Abstandsmaß zur Berechnung des Abstandes zweier diskreter Verteilungen herangezogen, so entsprechen die erwähnten Objekte den Funktionen der empirischen bzw. hypothetischen Verteilung. Die Größe entspricht also dem maximalen Wert des absoluten Abstands zwischen empirischer und hypothetischer Klassenhäufigkeiten.¹⁶³

Zieht man auch hier das bereits in den vorherigen Abschnitten verwendete Beispiel heran, so wird im Falle der 21 Nullen des Roulettes A und der 18 Nullen des Roulettes B bei 185 Spielrunden schnell ersichtlich, dass der Wert der Tschebyscheff-Distanz auch hier bei Roulette A

¹⁶² Vgl. Nigrini, Mark, Digital Analysis Using Benford's Law, Vancouver, 2000, S. 118f.

¹⁶³ Vgl. Götze, Wolfgang/ Deutschmann, Christel/ Link, Heike, Statistik, München, 2002, S. 331.

höher ausfallen muss, da der Abstand zu den eigentlich zu erwartenden 5 Kasinogewinnen höher ausfällt. Unter der Annahme, dass alle anderen Ziffern mit der gleichen Häufigkeit auftreten, liegt der Wert für Roulette A bei $TD = 0,0864$ und für Roulette B bei $TD = 0,0703$. Hierbei ist jeweils die Klasse der Zahl null diejenige, bei welcher der größte Abstand zu der unterstellten Gleichverteilung auftritt.

3.4.10 Verzerrungsfaktor-Modell

Das Verzerrungsfaktormodell wurde von *Nigrini* entwickelt und vergleicht die Mittelwerte des zu untersuchenden Datensatzes mit denen der hypothetischen Verteilung. Im Allgemeinen hängt der Mittelwert jedoch nicht ausschließlich von der Verteilung, sondern auch von der Größe der zu analysierenden Werte ab. Demzufolge ist es hierbei zunächst notwendig, diese in einen einheitlichen Wertebereich zu transformieren: Das Komma einer jeden Zahl wird so verschoben, dass alle Werte auf dem Intervall $[10, 100)$ liegen.¹⁶⁴ Mathematisch umsetzbar ist dies mit der Formel

$$x^* = \frac{10 \cdot x}{10^{\text{int}(\log(x))}} \quad (18)$$

Der Verzerrungsfaktor berechnet sich mittels Formel (19) wie folgt:

$$VF = \frac{TM - EM}{EM} \quad (19)$$

Hierbei bezeichnet TM den tatsächlich zu beobachtenden Mittelwert und EM den erwarteten Mittelwert. Unterstellt man z.B. eine Gleichverteilung, so folgt für $EM = 0,5 \cdot (100 + 10) = 55$. Für die Benford-Verteilung ist die Berechnung etwas komplizierter. Dort kann der erwartete Mittelwert mittels

$$EM = \frac{90}{n \cdot \left(10^{\frac{1}{n}} - 1\right)} \quad (20)$$

berechnet werden. Zu erkennen ist, dass dieser abhängig von dem Strichprobenumfang n ist. Dies ist zwar für die Ergebnisse nicht unproblematisch, jedoch ist dieser Effekt für $n > 500$

¹⁶⁴ Vgl. Nigrini, Mark, *Digital Analysis Using Benford's Law*, Vancouver, 2000, S. 62.

¹⁶⁵ Vgl. Nigrini, Mark, *Digital Analysis Using Benford's Law*, Vancouver, 2000, S. 64.

vernachlässigbar klein. Der tatsächlich beobachtete Mittelwert der zu analysierenden Daten ergibt sich dann zu

$$TM = \frac{1}{n} \cdot \sum_{i=1}^n x_i^* \quad (21)$$

Die Größe des Verzerrungsfaktors gibt im Anschluss an, wie stark die relative Abweichung von dem zu erwartendem Mittelwert ist. Für $VF < 0$ bedeutet dies folglich, dass mehr kleinere Zahlen in dem zu analysierendem Datensatz sind als dies zu erwarten war. Im Falle von $VF > 0$ befinden sich entsprechend mehr größere Zahlen in der Stichprobe. Multipliziert man den Verzerrungsfaktor mit 100 und nimmt den Betrag von diesem Wert, so erhält man die prozentuale Abweichung von dem Mittelwert der hypothetischen Verteilung.¹⁶⁶

Das Verzerrungsfaktormodell eignet sich v.a. für den Vergleich mehrerer Stichproben hinsichtlich der Konformität mit der hypothetischen Verteilung. Je größer der Betrag des errechneten Wertes ausfällt, desto größer ist auch die Wahrscheinlichkeit, dass eine Abweichung von dieser vorliegt. Erfolgt der Einsatz im Rahmen der digitalen Ziffernanalyse, steigt somit auch die Wahrscheinlichkeit an, dass Manipulationen vorliegen, welche das beobachtete Ergebnis hervorrufen.

Führt man auch hier den Vergleich der beiden Roulette-Tische durch, so ergibt sich aufgrund der angenommenen Gleichverteilung eines nicht manipulierten Roulettes der erwartete Mittelwert von $EM = 29,16$. Roulette A, das bei 185 Runden 21 Nullen hervorgebracht hat, erzielt bei einem Mittelwert von $TM = 17,26$ einen Verzerrungsfaktor von $VF = -0,4081$, was zum einen darauf schließen lässt, dass mehr kleine Zahlen gefallen sind als eigentlich unterstellt wurde und der Mittelwert um 40,81% von dem hypothetischen Mittelwert abweicht. Roulette B mit 18 Nullen erzielt einen Mittelwert von $TM = 17,3$ und somit einen Verzerrungsfaktor von $VF = -0,4067$. Demzufolge liegt der empirische Mittelwert hierbei 40,67% unter dem erwarteten Mittelwert, was insgesamt eine bessere Annäherung an die Gleichverteilung vermuten lässt.

¹⁶⁶ Vgl. Nigrini, Mark, Digital Analysis Using Benford's Law, Vancouver, 2000, S. 66.

4. Prüfung der Eignungsqualität der Anpassungstests und statistischen Kennzahlen zur Bestimmung der Abweichungen von Benford's Law

4.1 Beschreibung der Methodik und Vorgehensweise

Um eine Aussage über die Eignung der vorgestellten statistischen Testverfahren und Maßzahlen zur Anwendung im Rahmen des Benford-Testes zu erhalten, ist es notwendig zu prüfen, inwieweit diese in der Lage sind, Manipulationen innerhalb eines Datensatzes aufzuspüren und zu erkennen. Insbesondere gilt es hierbei unterschiedliche Eigenschaften eines Datensatzes, wie z.B. die Größe des Stichprobenumfanges sowie die Anzahl der enthaltenen manipulierten Daten, zu berücksichtigen.

Ausgangspunkt der Untersuchung sind demzufolge Datensätze, die der Benford-Verteilung folgen bzw. eine möglichst starke Annäherung an diese aufweisen. Eine Möglichkeit, Datensätze bestimmter Verteilungen zu generieren, bietet die Monte-Carlo-Simulation¹⁶⁷. Mittels dieser numerischen Simulation werden demnach vierstellige Zufallszahlen auf dem Intervall [1.000; 9.999] erstellt, die eine möglichst hohe Konformität mit der Benford-Verteilung aufweisen. Da für jedes der Verfahren lediglich die einzelnen Stellen von Bedeutung sind oder die Zahlen ohnehin in einen einheitlichen Wertebereich transformiert werden, spielt die Zehnerpotenz der Zahlen dem Augenschein nach keine Rolle. Mittels der vierstelligen Zahlen wäre es theoretisch möglich, die digitale Ziffernanalyse auf die ersten vier Stellen anzuwenden. Da die Daten in einem einheitlichen Wertebereich liegen, sind weitere Transformationen nicht notwendig. Die Zahlen werden hierbei in Datensätzen mit den Stichprobenumfängen $n_1 = 150, n_2 = 300, n_3 = 500, n_4 = 1.000, n_5 = 5.000, n_6 = 10.000, n_7 = 100.000$ und $n_8 = 1.000.000$ zusammengefügt.

In einem zweiten Schritt gilt es, Manipulationen in den Datensätzen zu simulieren. In Anlehnung an die Untersuchungsergebnisse von *Hill* und *Watrin et al.* wird hierbei unterstellt, dass manipulierte Zahlen eine starke Annäherung an die Gleichverteilung aufweisen.¹⁶⁸ Aus diesem Grund wird auch hierbei auf die Monte-Carlo-Simulation zurückgegriffen und Datensätze generiert, die gleichverteilte vierstellige Zufallszahlen enthalten. Die enthaltenen Zahlen werden dazu verwendet, Zahlen der Benford-verteilten Datensätze zufallsbasiert zu ersetzen. Hierbei

¹⁶⁷ Diese wird in Abschnitt 4.2 detailliert erläutert.

¹⁶⁸ Vgl. Hill, Theodore P., Random-Number guessing and the first digit phenomenon, in: *Psychological Reports*, 1988, S. 969 und Watrin, Cristoph/ Struffert, Ralf, Benford's Law und Chi-Quadrat-Test – Chancen und Risiken des Einsatzes bei steuerlichen Prüfungen – Zugleich Anmerkungen zum Urteil des FG Münster vom 07.12.2005 I K 6384/03 E, in: *Der Betrieb*, 2006, S. 1749.

werden für die unterschiedlichen Stichprobenumfänge jeweils 0%, 1%, 5%, 10% und 20% des Zahlenmaterials substituiert und zeitgleich angenommen, dass die Datensätze somit in Höhe dieser relativen Anteile manipulierte Daten enthalten.

Im Anschluss werden für die dann insgesamt 40 vorliegenden Datensätze die in Abschnitt 3.4 erläuterten Testgrößen und Kennzahlen ermittelt. Untersuchungsgegenstand sind hierbei die ersten Ziffern, die zweiten Ziffern sowie die ersten beiden Ziffern in Kombination. Da die Ziffernverteilungen für höhere Stellen gegen eine Gleichverteilung konvergieren, welche jedoch ohnehin im Falle der manipulierten Zahlen unterstellt wird, wird insbesondere auf die Betrachtung der dritten und vierten Stelle verzichtet, da in der Konsequenz nicht davon ausgegangen werden kann, dass mittels der gewählten Vorgehensweise Abweichungen diagnostiziert würden. Sämtliche Anpassungstests werden zum Signifikanzniveau $\alpha = 0,05$ durchgeführt, wobei die zugehörige Nullhypothese H_0 „Der untersuchte Datensatz ist Benford-verteilt“ lautet. Im Falle der ersten Ziffern ergeben sich demnach neun Merkmalsklassen, da die Zahlen $d_1 = 1, \dots, 9$ auftreten können. Im Falle der zweiten Ziffern kann zusätzlich die Zahl null vorkommen, sodass hier zehn Merkmalsklassen auftreten. Für die gleichzeitige Betrachtung der ersten und zweiten Ziffern können die Zahlen $d_{12} = 10, 11, \dots, 99$ – also insgesamt 90 Merkmalsklassen – auftreten. Entsprechend werden die ersten beiden Ziffern aus den Zahlen herausgelesen und die Klassenhäufigkeiten ausgezählt, sodass diese mit den zu erwartenden Häufigkeiten verglichen und die Testgrößen und Kennzahlen berechnet werden können. Im Anschluss werden die Ergebnisse bzgl. der unterschiedlichen Datensätze verglichen und kritisch ausgewertet.

4.2 Die Monte-Carlo-Simulation

Als Monte-Carlo-Simulation bezeichnet man grundsätzlich eine Methode, mit welcher Zufallszahlen generiert bzw. für weitere Problemstellungen benutzt werden können.¹⁶⁹ Zuzuordnen ist diese somit den sog. „Stochastischen Simulationen“ innerhalb der „theoretischen Simulationen“.¹⁷⁰

¹⁶⁹ Vgl. Müller-Gronbach, Thomas/Novak, Erik/Ritter, Klaus, Monte Carlo-Algorithmen, Berlin Heidelberg 2012, S. 1 und Koxholt, Rolf, Die Simulation - ein Hilfsmittel der Unternehmensforschung, München/Wien 1967, S. 34 f.

¹⁷⁰ Vgl. Müller-Gronbach, Thomas/Novak, Erik/Ritter, Klaus, Monte Carlo-Algorithmen, Berlin Heidelberg 2012, S. 1.

Gemäß der VDI-Richtlinie 3633 versteht man unter einer Simulation das „Nachbilden eines Systems mit seinen dynamischen Prozessen in einem experimentierfähigen Modell.“ Die aus der Simulation gewonnenen Erkenntnisse können dann auf das reale System übertragen werden.¹⁷¹ Dies ist notwendig, wenn ein Experiment an einem realen System nicht oder nur unter äußerst schwierigen Bedingungen durchzuführen ist. Regelmäßig ist dies der Fall, wenn der Aufwand den Nutzen z.B. aufgrund zu hoher Kosten deutlich übersteigt.¹⁷² Wird das System durch ein mathematisches Modell charakterisiert, handelt es sich um eine theoretische Simulation. Im Gegensatz hierzu existieren des Weiteren experimentelle Simulationen, bei welchen ein spezielles Experiment materiell nachgeahmt wird.¹⁷³ Als Beispiel wird mittels eines Bremsprüfstandes einer Kfz-Werkstatt das Bremsverhalten eines PKWs experimentell simuliert. Wird ein mathematisches Modell zur Simulation des Wetters im Falle bestimmter Konstellationen von Hoch- und Tiefdruckgebieten herangezogen, handelt es sich hingegen um ein theoretisches Modell. Weiterhin erfolgt bei letzteren eine Unterscheidung hinsichtlich der Berücksichtigung von Zufallsereignissen. Finden diese keine Beachtung, erfolgen also alle Berechnungen unter Sicherheit, spricht man von „deterministischen Simulationen“. Diese können in der Regel durch Algorithmen der numerischen Mathematik eindeutig gelöst bzw. die Lösung näherungsweise bestimmt werden.¹⁷⁴ Findet der Zufall hingegen Berücksichtigung innerhalb der Simulation, spricht man von einer stochastischen Simulation. Als Ergebnis erhält man hierbei nicht nur eine Lösung, sondern vielmehr einen Bereich von Lösungen, der die Unsicherheit widerspiegelt. Da im Rahmen der stochastischen Simulationen häufig eine Vielzahl von Berechnungsschritten erforderlich ist, erfolgt die Durchführung in der Regel computergestützt.¹⁷⁵ Letzteres ist auch bei der Monte-Carlo-Simulation der Fall.

Ihren Namen erhielt die Monte-Carlo-Simulation von den Wissenschaftlern *Neumann* und *Ulam*, die den Namen „Monte-Carlo“ für ein Waffenprojekt während des zweiten Weltkrieges

¹⁷¹ Verband Deutscher Ingenieure, VDI-Richtlinie 3633, Blatt 1, 1993, S. 3.

¹⁷² Vgl. Kolonko, Michael, Stochastische Simulation – Grundlagen, Algorithmen und Anwendungen, Wiesbaden 2008, S. 1.

¹⁷³ Vgl. Hartmann, Stephan, Simulation, in: Mittelstraß, Jürgen [Hrsg.], Enzyklopädie Philosophie und Wissenschaftstheorie, Stuttgart 1995, Bd. 3, S. 807-809.

¹⁷⁴ Vgl. Kolonko, Michael, Stochastische Simulation – Grundlagen, Algorithmen und Anwendungen, Wiesbaden 2008, S. 1.

¹⁷⁵ Vgl. Koxholt, Rolf, Die Simulation - ein Hilfsmittel der Unternehmensforschung, München/Wien 1967, S. 9.

einsetzen. Der Ursprung ist hierbei in einem Casino des gleichnamigen Stadtteils in Monaco zu finden. Dort fungierte ein Roulette als einfacher Zufallszahlengenerator.¹⁷⁶

Hieraus lässt sich bereits erahnen, dass Zufallszahlen den Kern der Monte-Carlo-Simulation ausmachen. Mittels dieser werden die entsprechenden Unsicherheiten berücksichtigt.¹⁷⁷ Verallgemeinert spricht man auch von Zufallsvariablen, deren theoretische mathematische und statistische Grundlagen im Folgenden näher erläutert werden.

Eine formale Definition für den Begriff der Zufallsvariable findet sich u.a. bei *Eckstein*:

„Ist Ω die Ergebnismenge eines Zufallsexperiments, so heißt eine Funktion X , die jedem Ergebnis $\omega \in \Omega$ eine reelle Zahl $X(\omega)$ derart zuordnet, dass $X: \Omega \rightarrow \mathbb{R}$ und $\omega \rightarrow X(\omega)$ gilt, eine Zufallsgröße X (auf Ω).“¹⁷⁸

Die Funktionswerte $X(\omega) = x$ mit $\omega \in \Omega$ bezeichnen Realisierungen von X . Während Ω zu dem Wahrscheinlichkeitsraum (Ω, Σ, P) gehört, wobei P ein Wahrscheinlichkeitsmaß für die Ergebnisse und Σ die Potenzmenge - die Menge aller Teilmengen - darstellt, zählen die zugeordneten reellen Zahlen x zu dem Messraum (Ω', Σ') .¹⁷⁹ Einfach ausgedrückt bedeutet dies nichts anderes, als dass die Zufallsvariable X als Funktion jedem Element ω der Ergebnismenge eine reelle Zahl x zuordnet.¹⁸⁰ Verdeutlicht werden kann dies mittels eines herkömmlichen sechsseitigen Würfels. Bei diesem existieren sechs Ergebnisse $\omega_1 = \text{"Eins"}$, $\omega_2 = \text{"Zwei"}$, $\omega_3 = \text{"Drei"}$, $\omega_4 = \text{"Vier"}$, $\omega_5 = \text{"Fünf"}$ und $\omega_6 = \text{"Sechs"}$. Hierbei erfolgt die Zuordnung der Ergebnisse zu $X(\omega_1) = 1$, $X(\omega_2) = 2$, $X(\omega_3) = 3$, $X(\omega_4) = 4$, $X(\omega_5) = 5$ und $X(\omega_6) = 6$. Es ergibt sich die Ergebnismenge des Wahrscheinlichkeitsraumes zu

$$\Omega = \{(\omega_1), (\omega_2), (\omega_3), (\omega_4), (\omega_5), (\omega_6)\}$$

und die Potenzmenge zu

¹⁷⁶ Vgl. Frey, Herbert, Monte Carlo-Simulation – Quantitative Risikoanalyse für die Versicherungsindustrie, München 2011, S. 16 f.

¹⁷⁷ Vgl. Wolf, Klaus, Monte-Carlo-Simulation – Einsatz im Rahmen der Unternehmensplanung, in: Controlling 21/2009, S. 545.

¹⁷⁸ Eckstein, Peter, Statistik für Wirtschaftswissenschaftler, 2. Aufl., Wiesbaden 2010, S. 156.

¹⁷⁹ Müller-Gronbach, Thomas/Novak, Erik/Ritter, Klaus, Monte Carlo-Algorithmen, Berlin/Heidelberg 2012, S. 1 und Bankhofer, Udo/Vogel, Jürgen, Datenanalyse und Statistik, 1. Aufl., Wiesbaden 2008, S. 292.

¹⁸⁰ Vgl. Hedderich, Jürgen/Sachs, Lothar, Angewandte Statistik, 14. Aufl., Berlin/Heidelberg 2012, S. 178.

$$\Sigma = \left\{ \emptyset, (\omega_1), (\omega_2), \dots, (\omega_6), (\omega_1, \omega_2), \dots, (\omega_5, \omega_6), \right. \\ \left. (\omega_1, \omega_2, \omega_3), \dots, (\omega_4, \omega_5, \omega_6), \dots, (\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6) \right\}.$$

Durch P wird jedem möglichen Ereignis, also jeder Teilmenge des Wahrscheinlichkeitsraumes, eine Wahrscheinlichkeit zugeordnet. Es gilt hierbei unter der Annahme, dass der betrachtete Würfel ideal ist und somit alle Ergebnisse gleichverteilt sind:

$$P(\emptyset) = 0,$$

$$P(\omega_1) = P(\omega_2) = P(\omega_3) = P(\omega_4) = P(\omega_5) = P(\omega_6) = \frac{1}{6},$$

$$P(\omega_1, \omega_2) = \dots = P(\omega_5, \omega_6) = \frac{1}{3},$$

$$P(\omega_1, \omega_2, \omega_3) = \dots = P(\omega_4, \omega_5, \omega_6) = \frac{1}{2},$$

$$P(\omega_1, \omega_2, \omega_3, \omega_4) = \dots = P(\omega_3, \omega_4, \omega_5, \omega_6) = \frac{2}{3},$$

$$P(\omega_1, \omega_2, \omega_3, \omega_4, \omega_5) = \dots = P(\omega_2, \omega_3, \omega_4, \omega_5, \omega_6) = \frac{5}{6},$$

$$P(\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6) = 1.$$

Letzteres Ereignis wird auch als „sicheres Ereignis“, das Ereignis der leeren Menge \emptyset als „unmögliches Ereignis“ bezeichnet.

Die Ergebnis- und Potenzmenge im Messraum ergeben sich dann zu

$$\Omega' = \{1, 2, 3, 4, 5, 6\} \text{ und}$$

$$\Sigma' = \left\{ \emptyset, (1), (2), \dots, (6), (1, 2), \dots, (5, 6), \right. \\ \left. (1, 2, 3), \dots, (4, 5, 6), \dots, (1, 2, 3, 4, 5, 6) \right\}.$$

Die Beschreibung der Wahrscheinlichkeiten einer Zufallsvariable erfolgt mittels der Verteilungsfunktion sowie der Dichtefunktion bei stetigen und der Wahrscheinlichkeitsfunktion im Falle diskreter Zufallsvariablen. Betrachtet man zunächst den stetigen Fall, so gibt das Integral zwischen den Grenzen a und b über die Dichtefunktion die Wahrscheinlichkeit an, mit welcher die Zufallsvariable Werte auf dem Intervall $[a, b]$ annimmt. Es gilt also

$$P(a \leq X \leq b) = \int_a^b f(x) dx \quad (22)$$

für alle reelle Zahlen $a < b$. Analog hierzu gibt die Verteilungsfunktion $F(x)$ die Wahrscheinlichkeit an, dass die Zufallsvariable X Werte kleiner gleich x annimmt. Entsprechend gilt der Zusammenhang

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt. \quad (23)$$

Des Weiteren beträgt die Fläche unterhalb der Dichtefunktion stets den Wert eins. Hieraus folgt, dass für alle Verteilungsfunktionen gilt

$$\lim_{x \rightarrow \infty} F(x) = 1. \quad (24)$$

Bei diskreten Zufallsvariablen kann jeder Zufallsvariable eine direkte Wahrscheinlichkeit unmittelbar zugeordnet werden. Man spricht deshalb nicht von einer Dichtefunktion, sondern von einer Wahrscheinlichkeitsfunktion, die durch

$$\rho(x_i) = P(X = x_i) = p_i \quad (25)$$

charakterisiert wird.¹⁸³ Die Verteilungsfunktion berechnet sich hierbei durch das aufsummieren der Einzelwahrscheinlichkeiten zu

$$F(x) = P(X \leq x) = \sum_{i=1}^n \rho(x_i). \quad (26)$$

n gibt hierbei die Anzahl der Zufallsvariablen an.¹⁸⁴ Ein wichtiger Parameter, der Auskunft über die erwartete Lage der Zufallsvariable gibt, ist der Erwartungswert. Mit anderen Worten gibt dieser den arithmetischen Mittelwert der Zufallsvariablen gewichtet nach der Auftretenswahrscheinlichkeit an. Er berechnet sich für stetige Zufallsvariablen zu

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx \quad (27)$$

sowie mittels

$$E(X) = \mu = \sum_{i=1}^n x_i \cdot \rho(x_i) \quad (28)$$

¹⁸¹ Vgl. Ross, Sheldon M., Simulation, 5. Aufl., San Diego, 2013, S. 9f.

¹⁸² Vgl. Bankhofer, Udo/ Vogel, Jürgen, Datenanalyse und Statistik, 1. Aufl., Wiesbaden, 2008, S. 292.

¹⁸³ Vgl. Ross, Sheldon M., Simulation, 5. Aufl., San Diego, 2013, S. 9.

¹⁸⁴ Vgl. Bankhofer, Udo/ Vogel, Jürgen, Datenanalyse und Statistik, 1. Aufl., Wiesbaden, 2008, S. 292f.

im diskreten Fall.¹⁸⁵ Der Erwartungswert liefert jedoch keinerlei Aussagen dazu, wie stark die Werte der Zufallsvariablen von diesem Abweichen. Um die Deviation genauer zu beschreiben werden deshalb Streuungsparameter eingesetzt, zu deren wichtigsten Parametern die Varianz bzw. die Streuung gehört. Für stetige Zufallsvariablen gilt

$$\text{Var}(X) = \sigma^2 = \int_{-\infty}^{\infty} [x - E(X)]^2 \cdot f(x) dx \quad (29)$$

sowie für diskrete Zufallsvariablen

$$\text{Var}(X) = \sigma^2 = \sum_{i=1}^n [x_i - E(X)]^2 \cdot f(x_i).^{186} \quad (30)$$

Die Streuung stellt hierbei die Quadratwurzel der Varianz dar, es gilt also

$$\sigma = \sqrt{\sigma^2}.^{187} \quad (31)$$

Dem Ursprung nach erfolgt die Zufallszahlengenerierung mittels mechanischer Instrumente wie z.B. sich drehender Räder, eines Münzwurfes oder des Mischens von Karten. Die Monte-Carlo-Simulation greift jedoch auf die Computerunterstützung zurück. Grundsätzlich ist dies zunächst ein Widerspruch, da der Computer an und für sich mit deterministischen Modellen arbeitet und eine stochastische Erzeugung von Zahlen nicht wirklich möglich ist. Aus diesem Grund werden die vom Computer generierten Zufallszahlen auch als „Pseudozufallszahlen“ bezeichnet.¹⁸⁸ Mittels der heute gängigen Verfahren wird jedoch i.A. eine hinreichende Qualität der Zahlen erzielt, sodass die Verwendung von computergestützten Systemen nahezu keine Nachteile mit sich bringt.¹⁸⁹

Um Zufallszahlen bestimmter Verteilungen zu generieren, wird bei der Monte-Carlo-Simulation auf eine bestimmte Methodik zurückgegriffen, die sog. „Quantiltransformation“. Der Computer erstellt hierbei gleichverteilte Zufallsvariablen auf dem Intervall [0,1]. Somit kommt jede reelle Zahl zwischen null und eins exakt gleich häufig vor. Im Anschluss wird diese Zufallszahl von der Ordinate ausgehend auf die Verteilungsfunktion projiziert und der Wert der Abzisse

¹⁸⁵ Vgl. Ross, Sheldon M., Simulation, 5. Aufl., San Diego, 2013, S. 11f.

¹⁸⁶ Vgl. Bankhofer, Udo/ Vogel, Jürgen, Datenanalyse und Statistik, 1. Aufl., Wiesbaden, 2008, S. 294.

¹⁸⁷ Vgl. Degen, Horst/ Lorscheid, Peter, Statistik-Lehrbuch mit Wirtschafts- und Bevölkerungsstatistik, 2. Aufl., München/Wien, 2002, S. 260.

¹⁸⁸ Vgl. Ross, Sheldon M., Simulation, 5. Aufl., San Diego 2013, S. 39 und Koxholt, Rolf: Die Simulation - ein Hilfsmittel der Unternehmensforschung, München/Wien, 1967, S. 198 f.

¹⁸⁹ Vgl. Müller-Gronbach, Thomas/Novak, Erik/Ritter, Klaus, Monte Carlo-Algorithmen, 1. Aufl., Berlin/Heidelberg, 2012, S. 3.

durch vertikale Projektion ermittelt und für jede Zufallszahl wiederholt. Dies bedeutet nichts anderes, als dass die Zufallszahlen in die Umkehrfunktion der Verteilungsfunktion eingesetzt und ausgewertet werden. Es gilt also

$$x = F^{-1}(y), \quad (32)$$

wobei x die Zufallszahlen der gewünschten Verteilung und y die vom Computer generierten Zufallszahlen auf dem Intervall $[0,1]$ darstellen. Allgemeine Voraussetzung ist jedoch hierfür, dass die Umkehrfunktion der Verteilungsfunktion bekannt ist bzw. zumindest approximativ angenähert werden kann.¹⁹⁰

Je mehr Zufallszahlen erzeugt werden, desto größer ist im Allgemeinen die Wahrscheinlichkeit, dass die erzeugten Zahlen gegen die unterstellte Verteilung konvergieren. Mit anderen Worten weisen Zufallszahlen mit einem kleinen Stichprobenumfang eine geringere Annäherung an die zu Grunde gelegte Verteilung auf als dies im Falle großer Stichprobenumfänge der Fall ist. Dies wird auch als *Gesetz der Großen Zahlen* bezeichnet und findet ebenfalls im Falle der Monte-Carlo-Simulation Berücksichtigung. Dies bedeutet auch, dass die Genauigkeit der Simulation mit der Anzahl der Simulationsdurchläufe ansteigt. Veranschaulicht wird dieser Effekt in den Abbildungen 5 bis 8: Beispielhaft wurden hierzu standardnormalverteilte Zufallszahlen mit dem Erwartungswert $E(X) = 0$ und der Standardabweichung $\sigma = 1$ erstellt. Zunächst ist die Dichtefunktion in Abbildung 5 dargestellt, wie sie im Falle einer idealen Standardnormalverteilung auftritt.

¹⁹⁰ Vgl. Frey, Herbert C./ Nießen, Gero, Monte Carlo Simulation, München 2001, S. 100f.

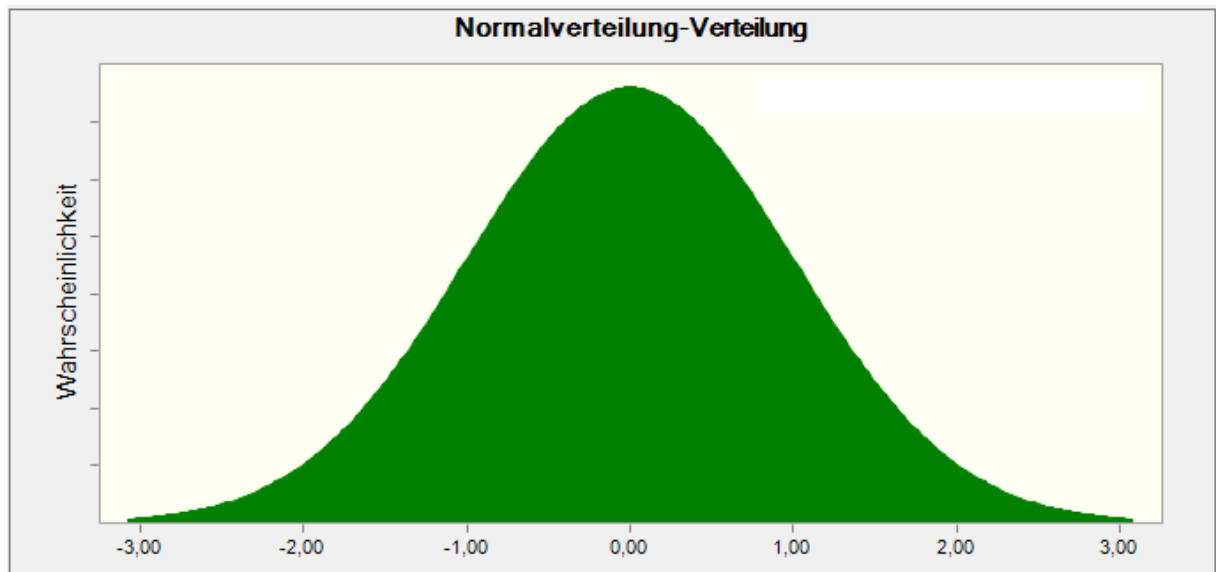


Abbildung 5: Dichtefunktion der Standardnormalverteilung mit Erwartungswert von null und Standardabweichung von eins

Abbildungen 6 bis 8 zeigen die Häufigkeitsverteilung von Zufallszahlen die mittels der Monte-Carlo-Simulation ermittelt wurden. In Abbildung 6 wurden zunächst $n = 500$ Simulationsläufe durchgeführt. Zu erkennen ist, dass die Verteilung deutlich von der idealen Standardnormalverteilung abweicht und somit nur eine begrenzte Annäherung zu beobachten ist.

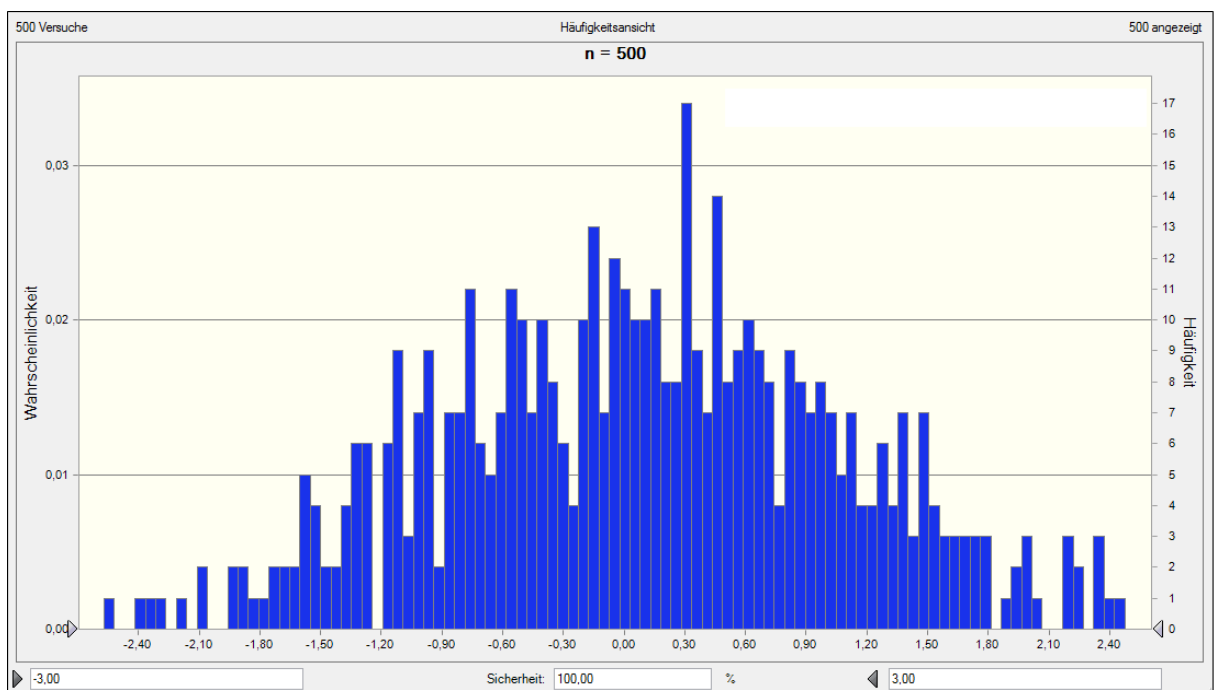


Abbildung 6: Häufigkeitsverteilung der mittels Monte-Carlo-Simulation erzeugten Zufallszahlen für $n = 500$

Abbildung 7 zeigt die Häufigkeitsverteilung im Fall $n = 10.000$ Simulationen. Hier ist bereits eine stärkere Ähnlichkeit zu der idealen Wahrscheinlichkeitsverteilung zu beobachten. Entsprechend liegt hier eine deutlich bessere Annäherung an die Standardnormalverteilung vor.

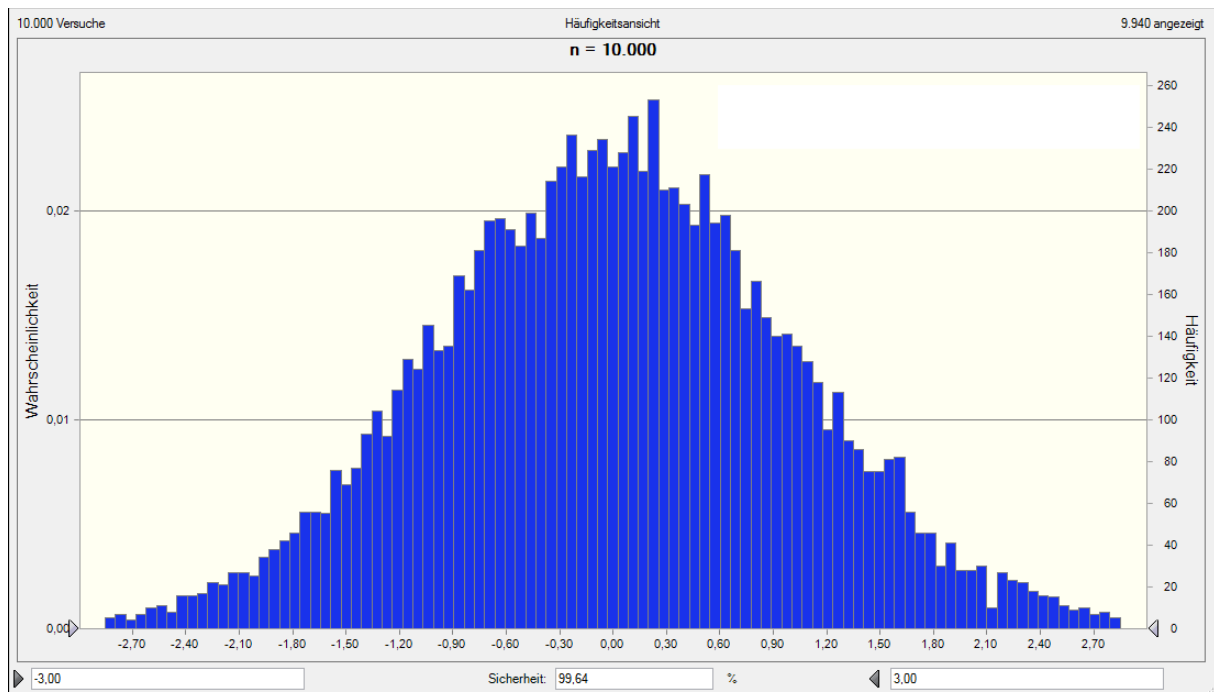


Abbildung 7: Häufigkeitsverteilung der mittels Monte-Carlo-Simulation erzeugten Zufallszahlen für $n = 10.000$

Abbildung 8 zeigt selbiges Experiment im Falle $n = 1.000.000$ Durchläufen. Hier entspricht die Häufigkeitsverteilung nahezu der idealen Standardnormalverteilung. Die Abweichungen von dieser sind in diesem Fall deutlich geringer.

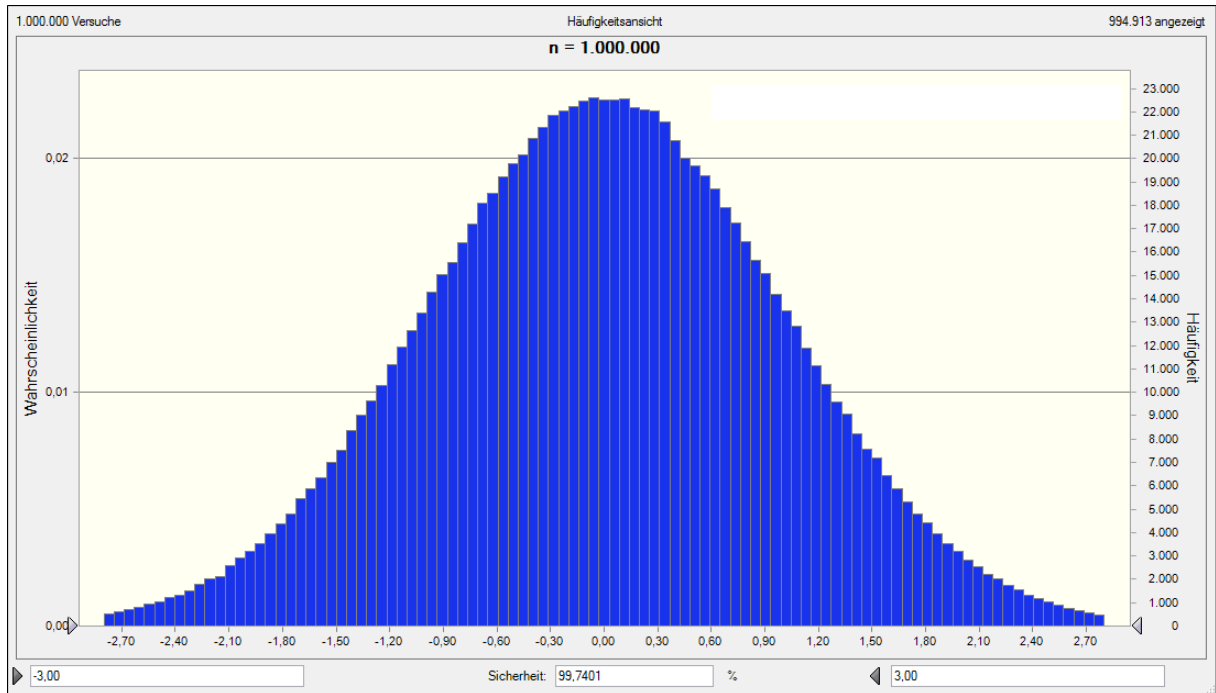


Abbildung 8: Häufigkeitsverteilung der mittels Monte-Carlo-Simulation erzeugten Zufallszahlen für $n = 1.000.000$

Wie groß die Anzahl der mindestens durchgeführten Simulationen sein muss, um aussagekräftige Ergebnisse zu erhalten ist im Allgemeinen nicht pauschal zu sagen. Vielmehr hängt dies von der Varianz der angenommenen Wahrscheinlichkeitsverteilung ab.¹⁹¹ Im Falle der Benford-Verteilung wird dieser Fragestellung in Kapitel 4.3 genauer nachgegangen. Jedoch bleibt festzuhalten, dass der Umstand des Gesetzes der Großen Zahlen im Rahmen der Datenauswertung in dem nächsten Abschnitt unweigerlich berücksichtigt werden muss.

4.3 Vergleich der Resultate der einzelnen Verfahren

Bevor eine genaue Auswertung in Anlehnung an den Benford-Test durchgeführt werden kann, ist es notwendig, die in Abschnitt 3.2 dargelegten Voraussetzungen zu überprüfen, um die Interpretierbarkeit der Ergebnisse sicherzustellen. Die Zufallszahlen wurden mit dem Microsoft Excel Add-In *Crystal Ball* des Softwareherstellers *Oracle* generiert, welche eine softwaretechnische Umsetzung der Monte-Carlo-Simulation darstellt. Hierzu wurde in dem Add-In eine benutzerdefinierte diskrete Verteilung erstellt, wobei die Wahrscheinlichkeiten der jeweiligen Ziffern an der entsprechenden Stelle bis auf die achte Nachkommastelle hinterlegt wurde. Da die Daten einer Monte-Carlo-Simulation unter der Vorgabe einer Benford-Verteilung entstammen, ist es selbstverständlich anzunehmen, dass sämtliche Bedingungen eingehalten werden.

¹⁹¹ Vgl. Frey, Herbert C./ Nießen, Gero, Monte Carlo Simulation, München, 2001, S. 100f.

Der zufällige Ursprung der Daten ist durch die Generierung von Zufallszahlen in jedem Fall sichergestellt. Da die Daten einheitenlos sind, stellen auch unterschiedliche Dimensionen keine Hindernisse dar. Zwar befinden sich alle Zahlen auf dem Intervall $[1000, 9999]$, jedoch stellen die Intervallgrenzen keine kritischen Ober- und Untergrenzen dar, da die Zahlen 1.000 und 9.999 jeweils die natürlichen Grenzen der vierten Zehnerpotenz im dekadischen Zahlensystem darstellen, was für die Benford-Verteilung unschädlich ist. Gut zu erkennen ist auch, dass sämtliche statistische Kennwerte, die in Tabelle 6 abgebildet sind, keineswegs gegen das Vorliegen einer Benford-Verteilung sprechen: Der Mittelwert ist im Falle aller Stichprobenumfänge größer als der Median und auch die Schiefe ist stets größer als null. Demzufolge spricht zunächst einmal nichts gegen die Durchführung eines Benford-Testes.

Stichproben- umfang n	Mittelwert	Median	Schiefe
150	3987,47	3281,00	0,72609
300	3756,29	3000,50	0,88873
500	3930,92	3091,50	0,81987
1.000	3904,87	3182,00	0,80415
5.000	3883,21	3104,00	0,79860
10.000	3943,38	3193,00	0,75037
100.000	3908,87	3161,00	0,78176
1.000.000	3728,17	2991,00	0,72380

Tabelle 6: Statistische Kennwerte der generierten Benford-verteilten Datensätze

Betrachtet werden zunächst die Häufigkeitsverteilungen der jeweiligen Ziffern derjenigen Datensätze, im Falle derer keine Substitutionen mit gleichverteilten Zahlen vorgenommen wurden. Diese repräsentieren die naturgemäß ideal Benford-verteilten Daten, die symbolisch frei von manipulierten Zahlen sind. Abbildung 9 stellt hierbei die Verteilungen der ersten Ziffern der sich hinsichtlich des Stichprobenumfanges unterscheidenden Daten dar. Gut zu erkennen ist, dass diejenigen Datensätze mit Stichprobenumfängen $n \geq 10.000$ Daten eine nahezu ideale Annäherung an die Benford-Verteilung aufweisen. Mit sinkendem Stichprobenumfang nehmen dann jedoch die Deviationen systematisch zu. Aufgrund des in Abschnitt 4.2 vorgestellten Gesetzes der großen Zahlen ist dies jedoch auch exakt in der Form zu erwarten: Mit steigendem Stichprobenumfang bzw. Zunahme der Anzahl an Simulationen erfolgt naturgemäß ein besserer

Fit der Daten an die unterstellte Verteilung. Demzufolge ist das Ergebnis der Monte-Carlo-Simulation als durchaus realitätsnah zu bewerten.

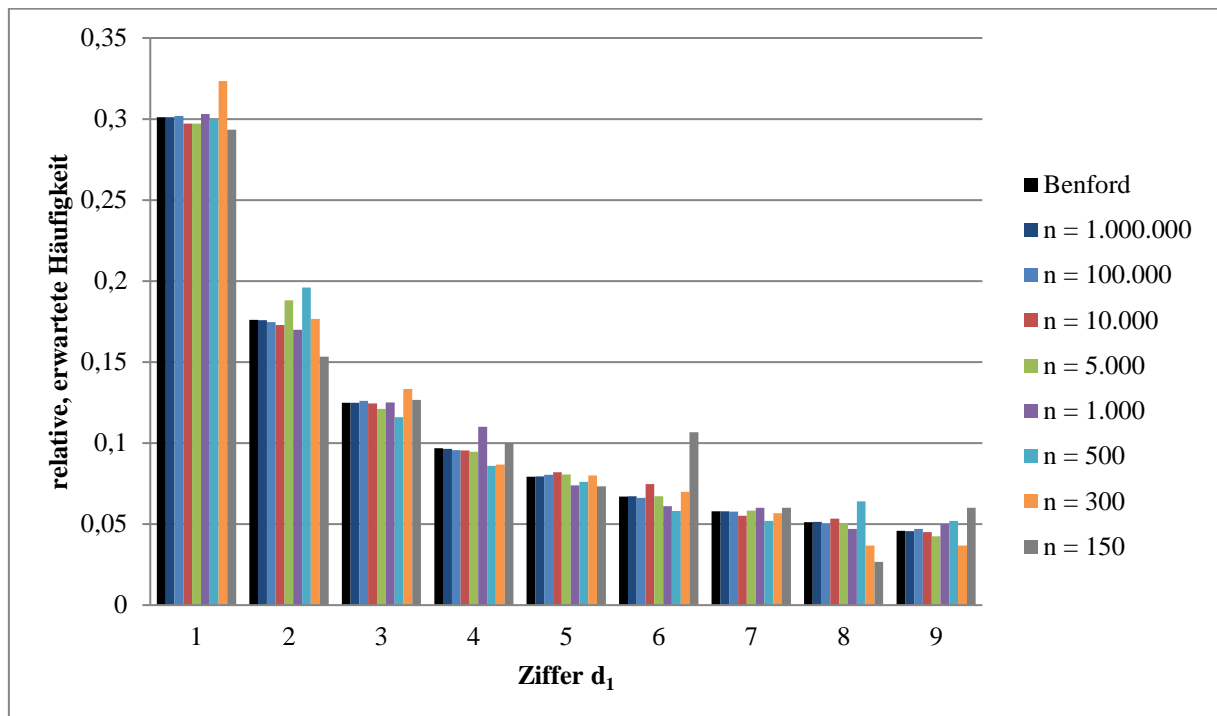


Abbildung 9: Verteilung der ersten Ziffern der Benford-verteilten Datensätze

Selbiges Phänomen kann auch für die Häufigkeitsverteilung der zweiten Ziffern in Abbildung 10 beobachtet werden: Während die Abweichungen bei den Datensätzen mit geringen Simulationszahlen auffällige Abweichungen aufweisen, folgen diejenigen mit großen Stichprobenumfängen nahezu exakt der Benford-Verteilung.

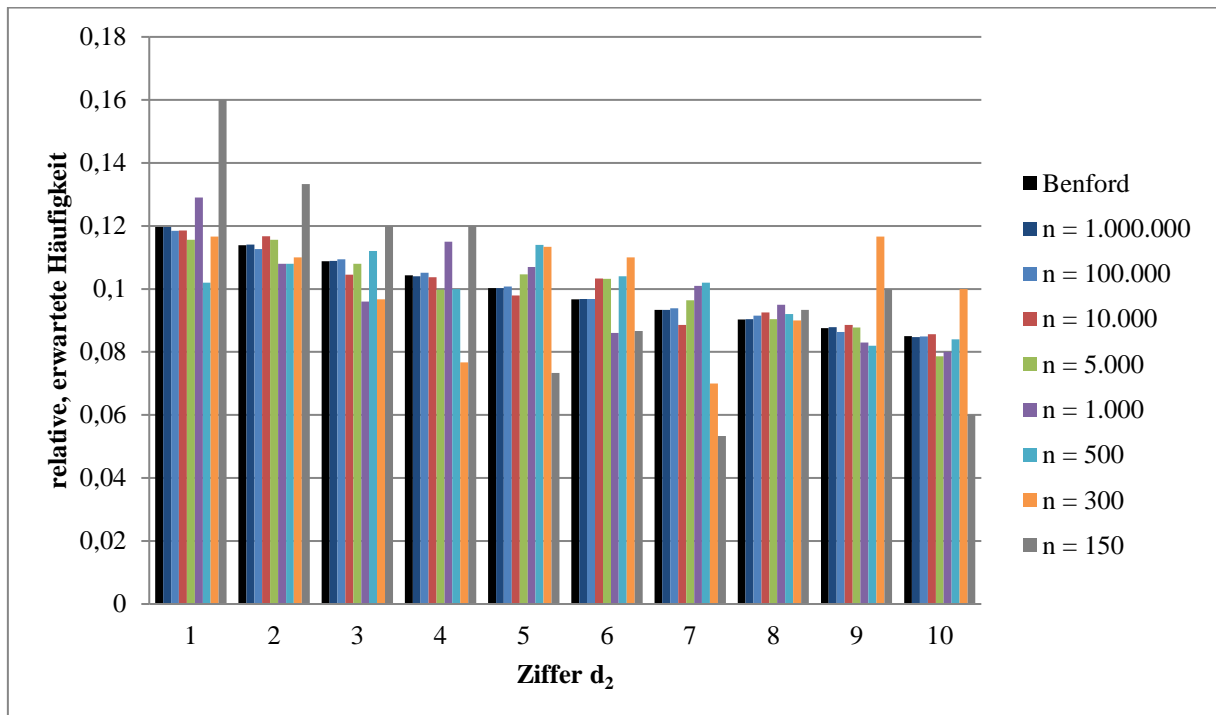


Abbildung 10: Verteilung der zweiten Ziffern der Benford-verteilten Datensätze

Für die in Kombination betrachteten ersten beiden Ziffern sind die Häufigkeitsverteilungen in den Abbildungen 11 bis 13 dargestellt. Abbildung 11 vergleicht hierbei die Häufigkeitsverteilung des kleinsten Datensatzes mit lediglich $n = 150$ Daten mit der Benford-Verteilung. Auffällig ist hierbei die relativ große Anzahl an Ausreißern, die entsprechend sowohl nach oben als auch nach unten auftreten.

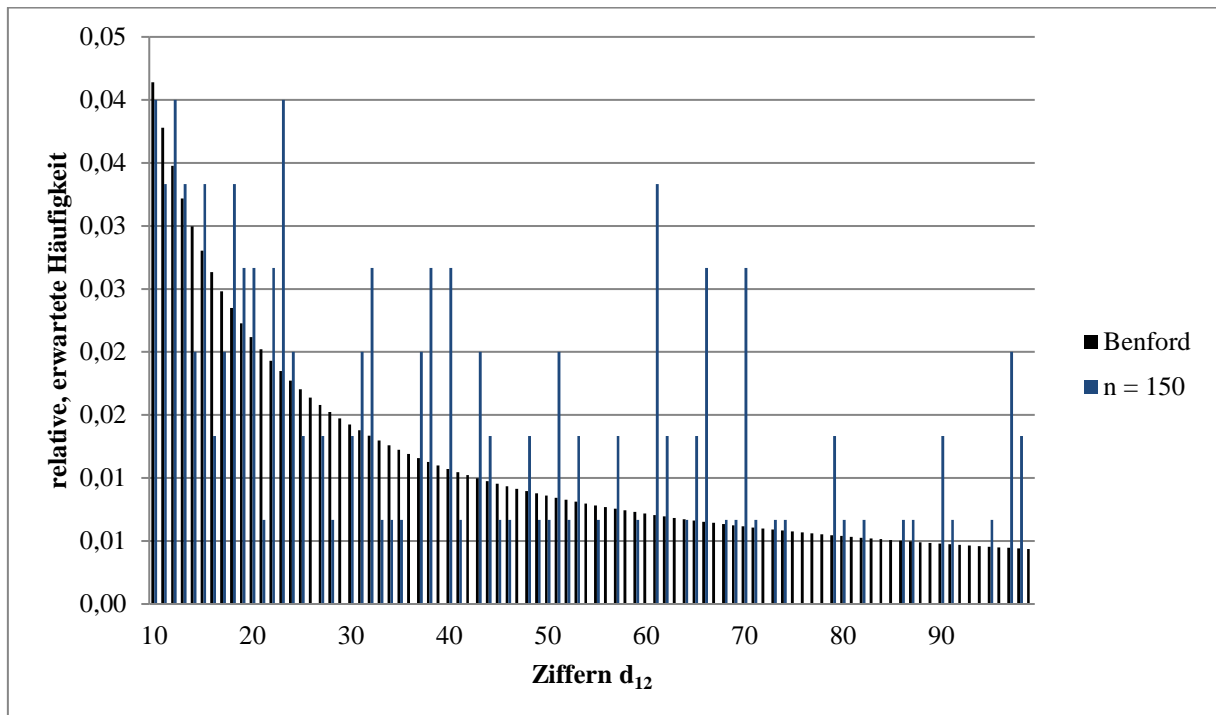


Abbildung 11: Häufigkeitsverteilung der ersten beiden Ziffern des Benford-verteilten Datensatzes mit $n = 150$ Daten

Abbildung 12 repräsentiert die Häufigkeitsverteilung des Benford-verteilten Datensatzes mit $n = 5.000$ Daten bzgl. der ersten beiden Ziffern. Treten hier vereinzelt noch Abweichungen auf, so ist hier doch insgesamt eine schon deutlich bessere Annäherung an die ideale Benford-Verteilung gegeben.

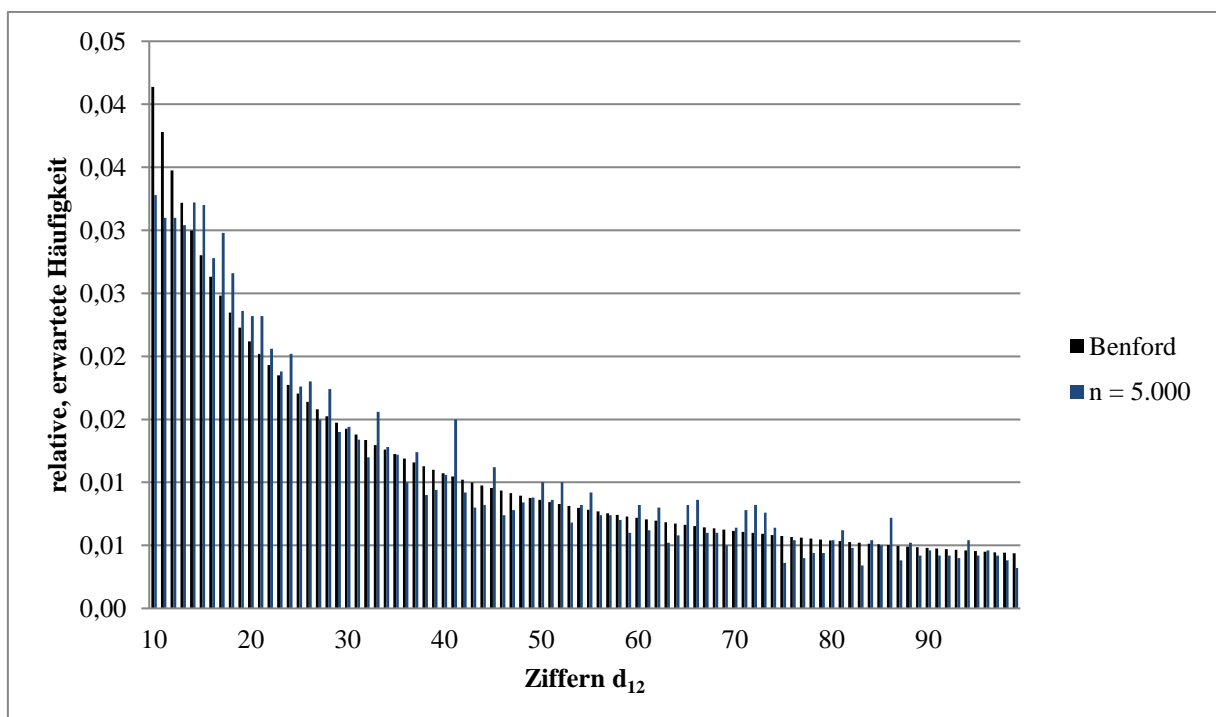


Abbildung 12: Häufigkeitsverteilung der ersten beiden Ziffern des Benford-verteilten Datensatzes mit $n = 5.000$ Daten

Die Häufigkeitsverteilung des größten Datensatzes mit $n = 1.000.000$ Daten stellt Abbildung 13 dar. Hier entspricht die Verteilung nahezu derjenigen, die theoretisch zu erwarten ist. Insgesamt kann deshalb auch bei Betrachtung der ersten beiden Ziffern das Gesetz der Großen Zahlen in seiner Form bestätigt werden. Fraglich ist nun, wie die einzelnen Testverfahren und Maßzahlen mit diesem Gesetz umgehen bzw. inwieweit diese theoretisch im Falle unterschiedlicher Stichprobenumfänge in der Lage sind, Manipulationen aufzudecken.

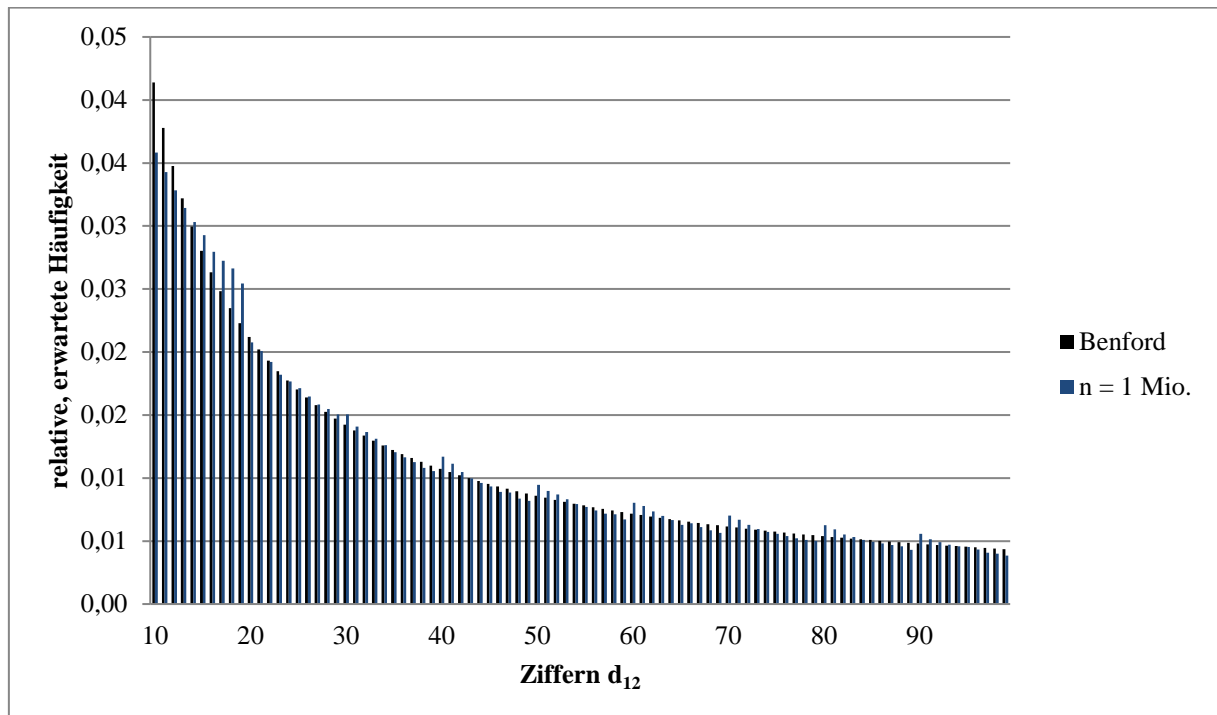


Abbildung 13: Häufigkeitsverteilung der ersten beiden Ziffern des Benford-verteilten Datensatzes mit $n = 1.000.000$ Daten

In einem weiteren Schritt wurden im Anschluss relative Anteile der Benford-verteilten Datensätze durch gleichverteilte Daten ersetzt, die wie in Abschnitt 4.1 beschrieben manipulierte Zahlen repräsentieren sollen. Hierbei wurden jeweils 0% (ideale Benford-Verteilung), 1%, 5%, 10% und 20% der Daten substituier und im Anschluss verschiedene statistische Testverfahren auf die Datensätze angewendet.

4.3.1 Ergebnisse des χ^2 -Anpassungstestes

Für die Durchführung des χ^2 -Anpassungstestes ist es zunächst einmal erforderlich, dass der Stichprobenumfang einen Mindestwert überschreitet, um signifikante Aussagen treffen zu können. Dieser begründet sich auf der minimalen Erwartungshäufigkeit der Klassen und divergiert somit für die Untersuchung der ersten, zweiten sowie der ersten beiden Ziffern. Gemäß der

herrschenden Meinung in der Literatur ist eine minimale Erwartungshäufigkeit von $E_i = 5$ pro Klasse erforderlich.¹⁹² Berechnet wird diese mittels

$$E_i = n \cdot p_i. \quad (33)$$

Für die ersten Ziffern ergibt sich somit ein minimaler Stichprobenumfang von 110^{193} , für die zweiten Ziffern von 59^{194} und für die ersten beiden Ziffern von 1.146^{195} . Für die Betrachtung der ersten und der zweiten Ziffern ergeben sich somit im Zuge der durchgeführten Untersuchung keinerlei Einschränkungen. Auf die Interpretation der Ergebnisse bzgl. der ersten beiden Ziffern muss jedoch bei den Stichprobenumfängen mit $n \leq 1.000$ verzichtet werden.

Der kritische Wert, bei dessen Überschreitung die Nullhypothese verworfen werden muss, bestimmt sich bei dem Chi²-Anpassungstest wie in Abschnitt 3.4.1 erläutert über das Quantil der Chi²-Verteilung, die in Anhang I verzeichnet ist. Dieses Quantil ist in erster Linie abhängig von den Freiheitsgraden der Untersuchung, die wiederum durch die Anzahl der vorhandenen Wahrscheinlichkeitsklassen charakterisiert sind. Dementsprechend unterscheidet sich auch der kritische Wert für die Untersuchungen der ersten, der zweiten sowie der ersten beiden Ziffern. Bei Analyse der ersten Ziffern treten insgesamt neun Merkmalsklassen auf, sodass der kritische Wert bei $X_{9-1;0,95}^2 = 15,5$ liegt. Im Falle der zweiten Ziffern existieren zehn Klassen, sodass die Nullhypothese abgelehnt werden muss, wenn die Testgröße den Wert $X_{10-1;0,95}^2 = 16,9$ überschreitet. Werden die ersten beiden Ziffern analysiert, beträgt die Anzahl der Wahrscheinlichkeitsklassen 90, sodass sich der kritische Wert zu $X_{90-1;0,95}^2 = 112,0$ ergibt.

Tabelle 7 gibt die Ergebnisse der Analyse bzgl. der ersten Ziffern wieder. Die hervorgehobenen Werte signalisieren jeweils eine signifikante Ablehnung der Nullhypothese. Zunächst einmal ist zu erkennen, dass es im Falle der idealen Benford-Verteilung zu keiner Ablehnung der Nullhypothese kommt. Bei 1% manipulierten Daten wird bei $n = 10.000$ und $n = 1.000.000$ die Nullhypothese abgelehnt. Im Falle eines Anteils von 5% wird diese zusätzlich bei $n = 100.000$

¹⁹² Vgl. Bomsdorf, Eckart, Wahrscheinlichkeitsrechnung und statistische Inferenz, 8. Aufl., Köln, 2002, S. 127.

¹⁹³ Die geringste Klassenwahrscheinlichkeit liegt bei der ersten Ziffern bei der Zahl 9 mit $p_9 = 0,0457575$. Es gilt: $\frac{5}{p_9} = \frac{5}{0,0457575} = 109,27 \leq n$.

¹⁹⁴ Die geringste Klassenwahrscheinlichkeit liegt bei der zweiten Ziffern bei der Zahl 9 mit $p_9 = 0,0849735$. Es gilt: $\frac{5}{p_9} = \frac{5}{0,0849735} = 58,83 \leq n$.

¹⁹⁵ Die geringste Klassenwahrscheinlichkeit liegt bei den ersten beiden Ziffern bei der Zahl 99 mit $p_{99} = 0,0043649$. Es gilt: $\frac{5}{p_{99}} = \frac{5}{0,0043649} = 1146 \leq n$.

abgelehnt, bei 10% und 20% für alle Stichprobenumfänge $n > 1.000$. Für $n \leq 1.000$ kommt es hingegen selbst bei 20% manipulierten Daten nicht zu einer Ablehnung. Auch fällt auf, dass je größer n ist, stets kleinere Anteile an manipulierten Daten ausreichen, um einen exorbitanten Anstieg des χ^2 -Wertes zu bewirken. Bei dem größten Datensatz mit lediglich 5% gleichverteilten Daten beträgt die χ^2 -Größe bereits 1.040,63. Bei $n = 10.000$ liegt dieser Wert bei gerade einmal 37,37, für $n = 300$ lediglich bei 4,03.

1. Ziffern	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
0%	6,5223	2,8162	5,0716	3,6590	6,5025	13,5459	12,8597	5,3507
1%	7,3235	2,9786	5,3255	3,8484	7,1629	15,6965	13,0019	51,5915
5%	7,0805	4,0290	5,8269	8,0147	10,0798	37,3679	116,5101	1.040,6293
10%	7,3719	5,2208	7,0971	7,7389	23,0479	81,4973	403,5202	4.133,8625
20%	11,9257	8,0548	8,8534	13,2982	96,6058	210,1502	1.647,4623	16.094,608

Tabelle 7: Ergebnisse des χ^2 -Anpassungstestes bei Betrachtung der ersten Ziffern

Tabelle 8 stellt die Ergebnisse des χ^2 -Anpassungstestes für die zweiten Ziffern dar. Hier fällt unmittelbar auf, dass die Nullhypothese für $n \leq 5.000$ nie abgelehnt wird. Bei $n = 10.000$ tritt die Ablehnung bei einem Anteil von 20% manipulierter Daten mit einem Wert von $X^2 = 21,53$ auf. Im Falle des Stichprobenumfanges von $n = 100.000$ Daten kommt es bei 10% simulierter Manipulationen zu einer Rejektion der Nullhypothese. Hier beträgt $X^2 = 24,85$. Sind 20% der Daten manipuliert steigt der χ^2 -Wert auf 63,39. Bei $n = 1.000.000$ erfolgt die signifikante Ablehnung der Benford-Verteilung bei 5% gefälschten Daten mit $X^2 = 37,13$. Beträgt deren Anteil 20% so liegt die Größe bei $X^2 = 472,87$. Ansonsten kann die gleiche Beobachtung wie bei Analyse der ersten Ziffern gemacht werden: Mit ansteigendem Stichprobenumfang genügen geringere Anteile fraudulenter Daten, um die Nullhypothese abzulehnen. Jedoch ist dieser Effekt bei der zweiten Ziffern nicht so deutlich erkennen.

2. Ziffern	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
0%	8,2622	9,1792	3,4025	6,6523	7,8572	10,7758	7,1125	3,8656
1%	8,2725	10,2102	3,7466	5,5825	7,2632	11,3337	6,9621	4,5743
5%	7,7851	5,9538	3,7849	6,0807	7,3225	13,0678	13,5842	37,1309
10%	9,4419	7,0676	5,3250	6,2685	8,0811	13,4296	24,8484	128,7630
20%	9,4796	12,5251	4,9510	6,9294	8,4275	21,5314	63,3907	472,8699

Tabelle 8: Ergebnisse des Chi²-Anpassungstestes bei Betrachtung der zweiten Ziffern

Die Ergebnisse der Analyse der ersten beiden Ziffern sind in Tabelle 9 dargestellt. Die kursiv dargestellten Werte lassen jeweils aufgrund des zu geringen Stichprobenumfanges keine Interpretationen zu. Für alle anderen Daten wird die Benford-Verteilung signifikant abgelehnt, selbst im Falle der eigentlich ideal verteilten Daten ohne jeglichen Manipulationsanteil. Hieraus lässt sich bereits folgern, dass die Anwendung des Chi²-Anpassungstestes bei Analyse der ersten beiden Ziffern nicht als geeignet erscheint.

1. u. 2. Zif.	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
0%	<i>108,274</i>	<i>81,350</i>	<i>72,823</i>	<i>111,908</i>	114,540	136,361	497,245	4.400,572
1%	<i>117,131</i>	<i>83,230</i>	<i>69,673</i>	<i>107,256</i>	114,358	142,788	497,777	4.320,135
5%	<i>111,745</i>	<i>77,422</i>	<i>68,951</i>	<i>120,407</i>	112,690	181,813	554,974	4.980,282
10%	<i>106,544</i>	<i>89,492</i>	<i>70,468</i>	<i>115,249</i>	125,383	214,837	811,079	7.746,984
20%	<i>144,890</i>	<i>151,274</i>	<i>70,568</i>	<i>120,712</i>	207,387	339,554	2.022,414	19.108,630

Tabelle 9: Ergebnisse des Chi²-Anpassungstestes bei Betrachtung der ersten beiden Ziffern

Insgesamt kann beobachtet werden, dass mit steigendem Anteil der gleichverteilten Daten auch die Werte der Testgrößen zunehmen, was selbstverständlich in der Form auch zu erwarten ist und die generelle Eignung des Chi²-Anpassungstestes im Rahmen des Benford-Testes bestätigt. Wie bereits oben beschrieben, fällt der Anstieg jedoch mit steigendem Stichprobenumfang sehr viel deutlicher aus. Zieht man des Weiteren die bereits dargestellten Häufigkeitsverteilungen in Betracht, so fällt ebenfalls auf, dass trotz der subjektiv vernommenen hohen Abweichungen der Datensätze mit niedrigen Stichprobenumfängen ohne substituierte Zahlen, die Chi²-Werte in einem sehr niedrigen Bereich von $X^2 \leq 10$ liegen. Beides lässt eindeutig darauf schließen, dass

das Gesetz der Großen Zahlen durch den Chi²-Anpassungstest sehr stark berücksichtigt wird: Der Theorie nach steigt die Anpassung der empirischen Daten an die unterstellte Verteilung mit steigendem Stichprobenumfang. Demzufolge sind Deviationen im Falle einer geringen Datenmenge natürlich, bei sehr großen Stichprobenumfängen sollten diese jedoch wesentlich geringer ausfallen. Bei kleinen Stichprobenumfängen können so jedoch auf Manipulationen zurückzuführende Deviationen deutlich schwieriger aufgedeckt werden. Dies hat jedoch auch zur Folge, dass die Ergebnisse unterschiedlicher Stichprobenumfänge nur eingeschränkt vergleichbar sind, was in den Ergebnistabellen eindeutig beobachtet werden kann: Im Falle von z.B. 10% fraudulent Daten liegt der Chi²-Wert bei den ersten Ziffern zwischen 5,22 und 4.133,86. Liegen die Stichproben jedoch in einem ähnlichen Bereich, sind die Ergebnisse auch bei kleineren Umfängen valide und vergleichbar. Lediglich auf die alleinige Interpretation statistisch signifikanter Entscheidungen sollte im Zweifel verzichtet werden, da vor allem bei geringen Stichprobenumfängen die Anteile fraudulenter Zahlen sehr hoch sein müssen, um eine Ablehnung der Nullhypothese zu implizieren.

Das diese Effekte bei Untersuchung der zweiten Ziffern deutlich geringer ausfallen ist damit zu begründen, dass die ideale Benford-Verteilung der zweiten Ziffern einer Gleichverteilung deutlich näher kommt. Da die simulierten manipulierten Daten jedoch eine genau solche aufweisen, wird es selbstverständlich schwieriger die Abweichungen aufzudecken. Eine Analyse der ersten beiden Ziffern in Kombination ist mit dem Chi²-Anpassungstest auf Basis der gewonnenen Informationen definitiv nicht empfehlenswert.

4.3.2 Ergebnisse des KS-Tests

Bevor der KS-Test durchgeführt werden kann, muss auch hier zunächst der kritische Wert berechnet werden, ab welchem die Nullhypothese „Der Datensatz ist Benford-verteilt“ abgelehnt werden kann. Gemäß der Formel aus Tabelle 2 in Abschnitt 3.4.2 ergeben sich in Abhängigkeit des Stichprobenumfanges unterschiedliche Werte. Für die vorliegenden Datensätze gibt Tabelle 10 die entsprechenden Grenzen wieder.

KS-Test	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
Kritischer Wert	0,0937	0,0663	0,0513	0,0363	0,0162	0,0115	0,0036	0,0011

Tabelle 10: kritische Werte des KS-Testes in Abhängigkeit der Stichprobenumfänge

Die Resultate der Analyse der ersten Ziffern sind in Tabelle 11 beschrieben. Auch hier kommt es bei den ideal Benford-verteilten Daten zu keiner Ablehnung der Nullhypothese. Diese treten ähnlich dem Chi²-Anpassungstest vorwiegend bei den großen Stichproben mit $n \geq 1.000$ auf. So erfolgt eine signifikante Ablehnung der Benford-Verteilung für die Daten mit einem Anteil manipulierter Zahlen von 1% bei $n = 10.000$ und $n = 1.000.000$. Bei den kleineren Datensätzen kommt es lediglich bei $n = 150$ Zahlen und einem Anteil von 20% gleichverteilter Daten zu einer Verwerfung. Betrachtet man die Testgrößen jedoch genauer, so lässt sich feststellen, dass – ausgenommen dem Datensatz mit $n = 150$ Daten – die Werte innerhalb der relativen Anteile in einer äquivalenten Größenordnung liegen. Im Falle eines Anteils von 5% liegen die Testgrößen zwischen 0,0136 und 0,0302, bei 10% zwischen 0,0242 und 0,0369 sowie bei 20% zwischen 0,0421 und 0,0607. Hieraus lassen sich zwei Schlüsse ziehen: Zum einen wird das Gesetz der Großen Zahlen im Falle des KS-Testes somit ausschließlich mittels der stichprobenabhängigen kritischen Werte berücksichtigt, die mit steigendem Stichprobenumfang systematisch abnehmen. Demzufolge genügen auch hier bei großen Stichprobenumfängen kleinere Abweichungen, die zu der Rejektion der Nullhypothese führen. Des Weiteren sind die Ergebnisse der Testgrößen untereinander durchaus vergleichbar, zumindest sofern der Stichprobenumfang größer gleich 300 ist.

1. Ziffern	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
0%	0,0315	0,0313	0,0191	0,0090	0,0081	0,0092	0,0013	0,0006
1%	0,0382	0,0256	0,0209	0,0081	0,0058	0,0115	0,0013	0,0031
5%	0,0505	0,0302	0,0169	0,0291	0,0136	0,0242	0,0136	0,0137
10%	0,0771	0,0269	0,0290	0,0361	0,0242	0,0369	0,0259	0,0273
20%	0,1038	0,0421	0,0430	0,0491	0,0576	0,0607	0,0533	0,0537

Tabelle 11: Ergebnisse des KS-Testes bei Betrachtung der ersten Ziffern

Tabelle 12 gibt die Ergebnisse bei Betrachtung der zweiten Ziffern wieder. Wirft man auch hier einen Blick auf die Fälle der signifikanten Ablehnungen so kann der gleiche Effekt wie bei dem Chi²-Anpassungstest beobachtet werden: Grundsätzlich gilt die Tatsache, dass bei großen Stichprobenumfängen kleine Abweichungen zu einer Verwerfung der Nullhypothese führen, wenngleich der Effekt weniger stark ausgeprägt ist als bei den ersten Ziffern. So kommt es bei $n = 1.000.000$ erst bei 5% manipulierten Daten zu einer signifikanten Ablehnung. Die Be-

gründung ist auch hier mit sehr großer Wahrscheinlichkeit in der höheren Ähnlichkeit der Benford-Verteilung der zweiten Stelle zur Gleichverteilung zu finden. Im Vergleich zu den Ergebnissen der ersten Ziffern ist eine Vergleichbarkeit der Testgrößen nur für ähnliche Stichprobengrößen (z.B. $n = 500$ bis $n = 5.000$) möglich. Weichen die Stichprobenumfänge stark voneinander ab, so erscheint ein unmittelbarer Vergleich der Werte weniger sinnvoll.

2. Ziffern	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
0%	0,0866	0,0467	0,0247	0,0096	0,0077	0,0056	0,0024	0,0003
1%	0,0933	0,0534	0,0256	0,0104	0,0066	0,0065	0,0024	0,0007
5%	0,0933	0,0404	0,0296	0,0264	0,0099	0,0080	0,0039	0,0027
10%	0,0999	0,0470	0,0367	0,0184	0,0119	0,0095	0,0062	0,0049
20%	0,1066	0,0704	0,0367	0,0164	0,0120	0,0156	0,0108	0,0094

Tabelle 12: Ergebnisse des KS-Testes bei Betrachtung der zweiten Ziffern

Die Analyseresultate der ersten beiden Ziffern sind Tabelle 13 zu entnehmen. Im Gegensatz zu dem Chi²-Anpassungstest scheint hier die Durchführung des Testes zumindest möglich zu sein. Signifikante Ablehnungen erfolgen hierbei lediglich bei den größten Datensätzen mit $n = 100.000$ und $n = 1.000.000$. Jedoch sind hier – wie auch bei Betrachtung der ersten Ziffern – die Testgrößen für $n > 150$ bei gleichen Anteilen fraudulenter Daten in einem ähnlichen Bereich und somit durchaus zwischen den Stichprobengrößen vergleichbar.

1. u. 2. Ziffern	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
0%	0,0201	0,0044	0,0097	0,0078	0,0023	0,0024	0,0009	0,0015
1%	0,0201	0,0044	0,0077	0,0078	0,0023	0,0025	0,0009	0,0012
5%	0,0177	0,0044	0,0023	0,0088	0,0016	0,0021	0,0012	0,0005
10%	0,0177	0,0054	0,0063	0,0077	0,0018	0,0034	0,0030	0,0026
20%	0,0311	0,0101	0,0097	0,0081	0,0057	0,0078	0,0078	0,0067

Tabelle 13: Ergebnisse des KS-Testes bei Betrachtung der ersten beiden Ziffern

Zusammenfassend bleibt festzuhalten, dass die Testgrößen des KS-Testes für die ersten Ziffern in einem hohen Maße stichprobeninvariant sind, sodass Vergleiche der Ergebnisse auch im

Fälle unterschiedlicher Stichprobengrößen grundsätzlich möglich sind. Als Einschränkung sollten jedoch nicht zu wenige Daten in die Untersuchung einbezogen werden. Ein Stichprobenumfang von $n \geq 300$ erscheint hierbei als sinnvoll. Für die zweiten Ziffern liegt immerhin eine eingeschränkte Stichprobeninvarianz vor. Statistisch signifikante Entscheidungen sollten mittels dieses Testverfahrens erst ab einem Umfang von ca. $n = 5.000$ Daten vorgenommen werden, um tatsächlich aussagekräftig zu sein. Bei alleiniger Verwendung dieser könnte es andernfalls sein, dass vorhandene Manipulationen schlichtweg nicht erkannt würden und eine Anwendung des Verfahrens zumindest fragwürdig erscheint. Einem Einbezug in eine vergleichende Betrachtung mehrerer Datensätze steht indes nichts im Wege.

4.3.3 Ergebnisse des Kuiper-Tests

Für die Durchführung des Kuiper-Testes zur Prüfung auf eine Abweichung von der Benford-Verteilung liegt der Wert der Grenze zur Ablehnung von H_0 für $\alpha = 0,05$ bei 1,321.

Tabelle 14 stellt die Ergebnisse bei Betrachtung der ersten Ziffern dar. Wie auch bei den vorherigen Testverfahren steigen die signifikanten Ablehnungen von H_0 mit ansteigenden Stichprobenumfängen an, sodass hier bereits kleinste Deviationen ursächlich für die Rejektionen sein können. Sind mehr als 5% der Daten innerhalb des Datensatzes fraudulent, so nehmen wie schon bei dem Chi²-Anpassungstest die Testgrößen der großen Stichprobenumfänge mit $n \geq 100.000$ stark zu. So beträgt der Wert der Testgröße bei 20% gleichverteilten Daten und $n = 1.000.000$ bereits 53,75. Zum Vergleich beträgt dieser Wert bei $n = 150$ lediglich 1,29. Insgesamt lässt sich erkennen, dass eine Vergleichbarkeit der Ergebnisse lediglich bei kleineren Stichprobenumfängen von $n \leq 1.000$ zulässig ist, da hier die Werte in einer ähnlichen Größenordnung liegen. Auch wenn die Güte des Testes mit steigendem Stichprobenumfang zunimmt, nimmt die Vergleichbarkeit der Werte hingegen ab.

1. Ziffern	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
0%	0,5183	0,5470	0,8553	0,4219	0,8440	1,0589	0,7063	0,7391
1%	0,5762	0,4484	0,8553	0,4479	0,8156	1,1488	0,7032	3,1385
5%	0,6462	0,5289	0,7201	0,9237	1,0480	2,4208	4,3091	13,6891
10%	0,9580	0,7701	0,6526	1,1462	1,7129	3,6928	8,1817	27,2842
20%	1,2893	1,0896	0,9680	1,5594	4,0799	6,0755	16,8632	53,7503

Tabelle 14: Ergebnisse des Kuiper-Testes bei Betrachtung der ersten Ziffern

Hinsichtlich einer statistisch signifikanten Entscheidung ergibt sich bei der Analyse der zweiten Ziffern ein äquivalentes Bild zu den Ergebnissen des Chi²-Anpassungstestes sowie des KS-Testes, was den dargestellten Ergebnissen in Tabelle 15 nachempfunden werden kann. Demzufolge kommt es bei einem Stichprobenumfang von $n = 1.000.000$ bei 5% gleichverteilter Daten zu einer signifikanten Ablehnung, bei $n = 10.000$ erfolgt die Ablehnung erst bei 20% manipulierter Daten, sodass der Effekt erhöhter signifikanter Rejektionen der Nullhypothese bei großen Stichprobenumfängen auch hier im Vergleich zu den ersten Ziffern leicht abgeschwächt auftritt. Im Gegensatz zu dem KS-Test kann hier keine Invarianz bzgl. des Stichprobenumfanges beobachtet werden, was auch für kleinere Stichprobenumfänge gilt. Zwar liegen die Werte alle in einem ähnlichen Bereich, jedoch weist die Testgröße der idealen Benford-Verteilung für $n = 150$ den Wert 1,08 auf, bei dem Datensatz bestehend aus 5.000 Zahlen und 20% fraudulenter Daten beträgt dieser Wert jedoch gerade einmal 0,99. Ein Vergleich dieser Ergebnisgrößen hätte somit unweigerlich falsche Schlüsse zur Folge. Ein Ergebnisvergleich bei ungefähr gleichen Stichprobenumfängen steht indes nichts entgegen.

2. Ziffern	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
0%	1,0759	0,8171	0,7048	0,6026	1,0005	0,7371	1,1601	0,4611
1%	1,1587	0,9393	0,6417	0,6344	0,9275	0,9074	1,1506	0,7881
5%	1,1587	0,7068	0,7690	1,0476	1,0289	0,8673	1,2373	2,6540
10%	1,2415	0,8516	0,8272	0,9522	1,1281	0,9544	1,9634	4,8864
20%	1,3244	1,3180	0,8272	0,9204	0,9913	1,5653	3,4061	9,3841

Tabelle 15: Ergebnisse des Kuiper-Testes bei Betrachtung der zweiten Ziffern

Tabelle 16 gibt die Ergebnisse der Analyse der ersten beiden Ziffern wieder. Grundsätzlich zeigt sich hier bei sehr großen Stichprobenumfängen von $n \geq 5.000$ ein ähnliches Bild wie bei dem χ^2 -Anpassungstest: Selbst die ideal Benford-verteilten Daten führen zu signifikanten Verwerfungen der Benford-Verteilung. Entsprechend kann der Einsatz in Form des Signifikanztests lediglich für Stichprobengrößen von $n \leq 1.000$ erfolgen. Problematisch ist, dass in diesem Bereich die Deviationen aufgrund des Gesetzes der Großen Zahlen deutlich schlechter durch den Test aufgedeckt werden. Demzufolge scheidet die Eignung als Signifikanztest im Falle der Analyse der ersten beiden Ziffern aus. Einer Nutzung zum Vergleich zweier Stichproben steht im Grunde nichts entgegen: Mit kleineren Ausnahmen liegen alle Ergebnisse zumindest für $n \leq 1.000$ bei gleichen relativen Anteilen manipulierter Zahlen in einem zumindest ähnlichem Bereich.

1. u. 2. Ziffern	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
0%	0,8803	1,1681	0,9017	0,8235	2,2017	1,7146	4,5984	13,6412
1%	0,9632	1,1760	0,9017	0,8235	2,1733	1,8148	4,6111	13,8480
5%	1,0460	1,1299	0,7974	1,3377	1,9891	2,7568	5,2513	16,9672
10%	1,2116	1,1907	0,7703	1,4331	2,1752	3,8485	8,2835	27,2842
20%	1,3547	1,5988	1,0857	1,6363	4,3477	6,2251	17,0631	53,7503

Tabelle 16: Ergebnisse des Kuiper-Testes bei Betrachtung der ersten beiden Ziffern

Zusammenfassend bleibt festzuhalten, dass der Kuiper-Test grundsätzlich die Eignung besitzt, Deviationen von der Benford-Verteilung zu erkennen. Der Einsatz als Signifikanztest macht im Wesentlichen ausschließlich bei den ersten Ziffern für $n \geq 1.000$ und bei den zweiten Ziffern für $n \geq 10.000$ Sinn. Der Einsatz im Rahmen eines relativen Ergebnisvergleichs steht jedoch auch für kleinere Größen nichts im Wege. Eine reale Stichprobeninvarianz liegt jedoch wenn überhaupt nur eingeschränkt für kleine Stichprobenumfänge mit $n \leq 1.000$ Daten vor. Dies gilt für die erste, zweite und für die ersten beiden Ziffern in Kombination.

4.3.4 Ergebnisse der Z-Statistik

Im Gegensatz zu den klassischen Anpassungstests erfolgt die Berechnung der Testgröße im Falle der Z-Statistik für jede einzelne Klasse separat. In der ersten Hälfte der Tabellen 17 bis 19 sind die maximal auftretenden Z-Werte aufgeführt. Die untere Hälfte bildet die Anzahl der

Klassen innerhalb deren es zu einer Überschreitung des Maximalwertes kommt. Strenggenommen erfolgt die Rejektion einer Benford-Verteilung, sobald die Z-Größe bei einer Klasse den Wert von 1,96 überschreitet. Dass diese Vorgehensweise jedoch im Falle des Benford-Testes als wenig sinnvoll erscheint, zeigen die Ergebnisse bei Betrachtung der ersten Ziffern in Tabelle 17: Hier würden signifikante Verwerfungen auch bei denjenigen Daten vorgenommen, bei denen keinerlei gleichverteilte Daten hinzugefügt wurden. Auch lässt sich hierbei kein eindeutiges Muster erkennen, welche Eigenschaften der Stichprobe zu einer signifikanten Abweichung führen. Ein Vergleich der maximal auftretenden Z-Größen erscheint wenig aussagekräftig: Zum einen würden lediglich die Abweichungen einer Klasse berücksichtigt, zum anderen zeigt Tabelle 17, dass auf Grund der Heterogenität der Ergebnisse die Werte unter einander kaum zu vergleichen sind.

Eindeutiger erscheint hierbei die Betrachtung der Anzahl der Klassen, bei welchen der Z-Wert überschritten wird und für welche die Benford-Verteilung verworfen werden müsste. Hier zeigt sich, dass auch dann v.a. bei großen Stichprobenumfängen und bereits kleinen Anteilen manipulierter Daten, eine Vielzahl an Klassen den kritischen Wert übersteigt. Trifft man deshalb die Annahme, die Benford-Verteilung sei abzulehnen, wenn es in mehr als einer Klasse zu einer Verwerfung der Nullhypothese kommt, so erhält man für Stichprobenumfänge von $n \geq 5.000$ sinnvolle Ergebnisse. Eine Vergleichbarkeit der Resultate für unterschiedliche Stichprobengrößen scheint jedoch in der Form nicht möglich zu sein.

1. Ziffern	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
Max. Wert								
0%	1,7831	1,0078	1,2025	1,3458	2,1922	3,0421	1,9333	1,3065
1%	2,1098	1,0078	1,2025	1,4527	2,1179	3,0821	1,9182	3,9353
5%	2,1098	1,2699	1,2026	1,5596	2,0857	4,2025	6,4734	20,0793
10%	1,7831	1,0078	1,3537	1,2389	3,6272	4,9918	12,3022	40,8922
20%	2,0745	1,7544	1,8115	2,0358	6,8336	8,1238	25,9383	81,6918
Anzahl Überschreitungen								
0%	0	0	0	0	1	1	0	0
1%	1	0	0	0	1	1	0	6
5%	1	0	0	0	1	4	6	8
10%	0	0	0	0	3	5	7	9
20%	1	0	0	1	6	7	8	9

Tabelle 17: Ergebnisse der Z-Statistik bei Analyse der ersten Ziffern

Betrachtet man auch bei den zweiten Ziffern lediglich die Anzahl der signifikanten Abweichungen, so gelten auch für die Z-Statistik die gleichen Eigenschaften, die auch die anderen Testverfahren gezeigt haben: Der Effekt, dass kleine relative Abweichungen in großen Stichprobenumfängen für signifikante Deviationen sorgen, lässt sich hier in abgeschwächter Form beobachten, wie Tabelle 18 zu entnehmen ist. Die obige Vorgehensweise liefert im Falle der zweiten Ziffern nur für sehr große Stichproben ($n \geq 100.000$) plausible Ergebnisse. Ein Vergleich der Maximalwerte ist auch hier wenig aussagekräftig.

2. Ziffern	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
Max. Wert								
0%	1,5452	1,6808	1,1490	1,2512	1,5967	2,2242	1,3481	1,1443
1%	1,3956	1,8850	1,1490	1,4543	1,5460	2,2918	1,3270	1,5829
5%	1,3956	1,6808	1,0936	1,4543	1,6511	2,5287	2,0219	3,4683
10%	1,6492	1,6808	1,3913	1,4659	1,8254	2,5287	2,7168	6,0439
20%	1,5370	2,4961	1,1490	1,9831	2,0838	2,6979	4,0278	11,7496
Anzahl Überschrei- tungen								
0%	0	0	0	0	0	1	0	0
1%	0	0	0	0	0	1	0	0
5%	0	0	0	0	0	1	1	3
10%	0	0	0	0	0	1	4	7
20%	0	1	0	1	1	1	7	8

Tabelle 18: Ergebnisse der Z-Statistik bei Analyse der zweiten Ziffern

Da im Falle der Analyse der ersten beiden Ziffern deutlich mehr Merkmalsklassen vorhanden sind, steigt selbstverständlich auch die Anzahl der Ablehnungen im Rahmen der Z-Statistik an. Betrachtet man die Ergebnisse in Tabelle 19, so wird ersichtlich, dass die Höhe der Maximalwerte kaum von den Anteilen fraudulenter Daten abhängt. Selbiges lässt sich jedoch auch für die Anzahl der Abweichungen beobachten. Betrachtet man den Datensatz mit $n = 1.000.000$ Daten, so treten bei 0% Falschangaben bereits 66 Ablehnungen auf. Bei 10% manipulierten Daten steigt dieser Wert lediglich auf 69. Auch bei den kleineren Stichprobengrößen lässt sich ähnliches beobachten: Bei $n = 150$ Daten beträgt die Anzahl an Abweichungen des idealen Benford-Datensatzes 4 – der gleiche Wert ergibt sich wenn 10% der Daten manipuliert wurden. Demzufolge kann davon ausgegangen werden, dass die Z-Statistik sich für die Bestimmung der Abweichung von der Benford-Verteilung bei Analyse der ersten beiden Ziffern nicht eignet.

1. u. 2. Ziffern	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
Max. Wert								
0%	3,3549	1,8897	2,3516	2,4515	3,0814	3,5178	8,7912	27,9883
1%	3,5685	1,8897	2,3516	2,4515	3,3593	3,7365	8,8388	29,1179
5%	3,5685	1,8897	2,3844	2,8215	3,3593	3,9552	10,6962	34,1732
10%	3,3549	2,2886	2,3844	2,6037	3,6372	4,1075	12,1249	40,2677
20%	3,5685	3,6099	2,2506	2,8290	4,5056	5,9573	15,6968	52,2106
Anzahl Überschreitungen								
0%	4	0	1	6	5	9	29	66
1%	4	0	1	5	5	9	29	64
5%	3	0	3	7	5	13	26	60
10%	4	2	1	7	7	18	39	69
20%	6	6	1	6	15	24	62	80

Tabelle 19: Ergebnisse der Z-Statistik bei Analyse der ersten beiden Ziffern

Möchte man die Z-Statistik nutzen um zu prüfen, ob ein Datensatz der Benford-Verteilung folgt, ist es ratsam, die Anzahl der Klassen zu bestimmen, bei welchen es zu signifikanten Verwerfungen der Nullhypothese kommt. Es sollte hierbei davon ausgegangen werden, dass eine Ablehnung zu beobachten ist, wenn es zu mehr als einer Verwerfung kommt. Des Weiteren erscheint diese Vorgehensweise erst ab $n = 5.000$ Daten für die ersten Ziffern aussagekräftig zu sein. Bei den zweiten Ziffern sollten sogar $n = 100.000$ Werte vorliegen. Bei einer Betrachtung der ersten beiden Ziffern in Kombination ist diese Vorgehensweise nicht anwendbar.

Abgesehen von der Eignung zur Prüfung hinsichtlich der Konformität empirischer Daten mit der Benford-Verteilung ist es jedoch zusätzlich möglich, die Z-Statistik bei bereits attestierten Abweichungen heranzuziehen, um nachzuprüfen, welche Ziffern ursächlich für die Deviationen sind. Dies kann einfach durchgeführt werden, indem nachgegangen wird, in welchen Klassen es zu den größten Abweichungen kommt. Dies wäre grundsätzlich ein sinnvoller und empfehlenswerter Schritt, um Manipulationsmuster aufzudecken und so möglichen Deviationsursachen einen Schritt näher zu kommen.

4.3.5 Ergebnisse des normierten χ^2 -Wertes

Sowohl der normierte χ^2 -Wert als auch die weiteren folgenden Abstandsmaße stellen keine klassischen Anpassungstests der Statistik dar. Aus diesem Grund existieren keinerlei kritische Werte, bei deren Überschreitung eine unterstellte Verteilung abzulehnen wäre. Vielmehr können die entsprechenden Größen lediglich zwischen verschiedenen Untersuchungen verglichen werden, um die Untersuchungsobjekte zu identifizieren, welche die größte Abweichung von der Benford-Verteilung aufweisen.

Die Ergebnisse hinsichtlich der ersten Ziffern finden sich in Tabelle 20. Zunächst einmal ist zu erkennen, dass sämtliche Werte des Stichprobenumfanges von $n = 150$ Daten über denen der anderen Datensätze liegen. Dies lässt darauf schließen, dass der Stichprobenumfang für eine plausible Aussage bzgl. des normierten χ^2 -Wertes mindestens 300 betragen sollte. Ist dies gegeben, lässt sich jedoch eine sehr gute Vergleichbarkeit der berechneten Kennzahlen v.a. für ähnliche Stichprobenumfänge erkennen: Betrachtet man die Stichprobenumfänge zwischen $n = 300$ und $n = 1.000$, so liegen die Werte innerhalb gleicher relativer Anteile manipulierter Daten stets in einem ähnlichen Bereich. Sind keinerlei fraudulente Daten enthalten, ergeben sich Werte zwischen 0,0037 und 0,0094. Bei 5% gleichverteilter Zahlen liegen die Ergebnisse zwischen 0,0080 und 0,0134. Sind 20% der Daten manipuliert, ergibt sich das Ergebnisintervall zu $[0,0133; 0,0268]$. Ähnlich verhält es sich bei den großen Stichprobenumfängen von $n = 100.000$ und $n = 1.000.000$: Bei 1% gleichverteilten Daten liegen die Ergebnisgrößen zwischen 0,000052 und 0,000130, sind 20% der Daten fraudulent ergeben sich Ergebnisse zwischen 0,016095 und 0,016475. Die Vergleichbarkeit der Resultate wird lediglich durch deutliche Unterschiede in den Stichprobengrößen eingeschränkt. Vergleicht man die Umfänge mit $n = 300$ und $n = 1.000.000$ miteinander so stellt man fest, dass bei der idealen Verteilung des kleinen Datensatzes der normierte χ^2 -Wert bei 0,009387 liegt. Betrachtet man den großen Datensatz, so ergibt sich jedoch selbst bei 10% manipulierten Daten gerade einmal ein Wert von 0,004134. Hieraus lässt sich schließen, dass zwar faktisch eine Stichprobeninvarianz vorliegt, diese jedoch nur Gültigkeit besitzt, so lange die Stichprobenumfänge in zumindest ähnlichen Größenordnungen liegen und sich lediglich um den Faktor 10 unterscheiden.

1. Zif.	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
0%	0,043482	0,009387	0,010143	0,003659	0,001300	0,001355	0,000129	0,000005
1%	0,048823	0,009929	0,010651	0,003848	0,001433	0,001570	0,000130	0,000052
5%	0,047204	0,013430	0,011654	0,008015	0,002016	0,003737	0,001165	0,001041
10%	0,049146	0,017403	0,014194	0,007739	0,004610	0,008150	0,004035	0,004134
20%	0,079505	0,026849	0,017707	0,013298	0,019321	0,021015	0,016475	0,016095

Tabelle 20: Ergebnisse des normierten Chi²-Wertes bei Analyse der ersten Ziffern

Für die Ergebnisse bei Analyse der zweiten Ziffern, die in Tabelle 21 dargestellt sind, ergeben sich grundsätzlich die gleichen Beobachtungen wie im Falle der ausschließlichen Betrachtung der ersten Ziffern. Lediglich der Mindeststichprobenumfang, ab welchem interpretierbare Ergebnisse vorliegen, scheint hier bei $n = 500$ zu liegen. Wirft man einen Blick auf die Ergebnisse für $n \leq 300$ Daten, so liegen hier alle Werte in einem im Vergleich zu den restlichen Ergebnissen deutlich erhöhten Bereich. Demzufolge ist eine vergleichende Analyse der Resultate hier nicht möglich.

2. Zif.	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
0%	0,055081	0,030597	0,006805	0,006652	0,001571	0,001078	0,000071	0,000004
1%	0,055150	0,034034	0,007493	0,005583	0,001453	0,001133	0,000070	0,000005
5%	0,051900	0,019846	0,007570	0,006081	0,001464	0,001307	0,000136	0,000037
10%	0,062946	0,023559	0,010650	0,006268	0,001616	0,001343	0,000248	0,000129
20%	0,063198	0,041750	0,009902	0,006929	0,001685	0,002153	0,000634	0,000473

Tabelle 21: Ergebnisse des normierten Chi²-Wertes bei Analyse der zweiten Ziffern

Selbiges gilt auch hinsichtlich der Untersuchung der ersten beiden Ziffern. Die Ergebnisse sind in Tabelle 22 dargestellt. Zwar erscheint die Applikation der normierten Chi²-Größe für diesen Fall – im Gegensatz zu dem Chi²-Anpassungstest – als grundsätzlich möglich, jedoch sollte auch hier der Stichprobenumfang für den Erhalt aussagekräftiger Ergebnisse bei mindestens 500 liegen. Ist dies erfüllt, ist ein Vergleich Datensätze unterschiedlicher Stichprobengrößen möglich, solange diese nicht allzu stark voneinander divergieren.

1. u. 2. Z.	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
0%	0,721826	0,271167	0,145646	0,111908	0,022908	0,013636	0,004972	0,004401
1%	0,780873	0,277435	0,139345	0,107256	0,022872	0,014279	0,004978	0,004320
5%	0,744968	0,258072	0,137902	0,120407	0,022538	0,018181	0,005550	0,004980
10%	0,710291	0,298306	0,140936	0,115249	0,025077	0,021484	0,008111	0,007747
20%	0,965931	0,504246	0,141136	0,120712	0,041477	0,033955	0,020224	0,019109

Tabelle 22: Ergebnisse des normierten Chi²-Wertes bei Analyse der ersten beiden Ziffern

Allgemein lässt sich zusammenfassen, dass im Falle des normierten Chi²-Wertes durchaus eine Invarianz bzgl. des Stichprobenumfanges vorliegt. Diese unterliegt jedoch der Einschränkung, dass sich die Umfänge nicht allzu stark unterscheiden dürfen. So sind Vergleiche der Ergebnisse zwischen Datensätzen mit 300 und 1 Mio. Daten nicht möglich, wohl jedoch mit beispielsweise 300 und 10.000 Werten. Des Weiteren sollte im Falle jeder Untersuchung, also sowohl bei der ausschließlichen Betrachtung der ersten Ziffern, der zweiten Ziffern und der ersten beiden Ziffern, die Stichprobe stets eine Mindestgröße aufweisen. Sind diese Restriktionen eingehalten, eignet sich der normierte Chi²-Wert sehr gut, um unterschiedliche Untersuchungsobjekte hinsichtlich einer Benford-Deviation miteinander zu vergleichen.

4.3.6 Ergebnisse des normierten euklidischen Abstandes d*

Die Ergebnisse bei Betrachtung der ersten Ziffern stellt Tabelle 23 dar. Zu erkennen ist, dass v.a. im Falle der kleinen Stichprobenumfänge die Werte im Vergleich zu den übrigen Ergebnissen relativ groß ausfallen. Zu begründen ist dies damit, dass die normierte euklidische Abstandsmessung lediglich die relativen Abweichungen in die Berechnung einbezieht. Diese sind jedoch, auch bei den hypothetisch ideal Benford-verteilten Daten, aufgrund des Gesetzes der Großen Zahlen naturgemäß vorhanden. Aus diesem Grund werden auch bei dieser Testgröße aussagekräftige Schlüsse erst ab einem Stichprobenumfang von $n \geq 300$ zulässig. Des Weiteren ist auch hier die eingeschränkte Stichprobeninvarianz zu beobachten: Sehr gut vergleichbar sind Stichprobenumfänge in einer ähnlichen Größenordnung. Hier ergibt sich eine hervorragende Komparabilität der Ergebnisse. Vergleicht man jedoch die Resultate des Datensatzes mit 500 Zahlen mit demjenigen mit 1 Mio. Daten, so könnte man fälschlicherweise zu der Erkennt-

nis gelangen, dass ein Wert von 0,0294 bei dem kleinen Datensatz eine größere Deviation attestiert als der Wert 0,0226 bei dem großen Datensatz. Wirft man jedoch einen Blick auf den relativen Anteil fraudulenter Daten, bei denen diese Werte entstehen, fällt auf, dass genau Gegenteiliges der Fall ist: Der kleinere Datensatz weist hypothetisch keinerlei manipulierte Daten auf, wohingegen der Datensatz mit 1 Mio. Zahlen 10% gleichverteilte Zahlen enthält.

1. Zif.	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
0%	0,052924	0,030168	0,029371	0,017094	0,013284	0,010041	0,003189	0,000652
1%	0,058423	0,032160	0,030259	0,017887	0,014127	0,011274	0,003210	0,002353
5%	0,060578	0,031828	0,032175	0,027958	0,017384	0,019216	0,011404	0,011225
10%	0,071144	0,034623	0,038597	0,029374	0,026597	0,029726	0,021909	0,022560
20%	0,101079	0,045761	0,043003	0,038983	0,050919	0,048540	0,044978	0,044731

Tabelle 23: Ergebnisse des normierten euklidischen Abstandes bei Analyse der ersten Ziffern

Tabelle 24 visualisiert die Ergebnisse der Analyse der zweiten Ziffern. Insgesamt ist erkennbar, dass sich die Ergebniswerte der Datensätze mit divergierenden relativen Manipulationsanteilen bei gleichem Stichprobenumfang nur gering voneinander unterscheiden. Zwar steigen die Werte mit zunehmendem Anteil gleichverteilter Zahlen an, jedoch nur in einem sehr geringen Maße. Aus diesem Grund erscheint es als sehr schwierig, Daten mit unterschiedlichen Stichprobenumfängen miteinander zu vergleichen. Divergieren die Stichprobenumfänge jedoch nur eingeschränkt, ist eine Anwendung durchaus realisierbar.

2. Zif.	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
0%	0,077505	0,055830	0,027928	0,027106	0,012770	0,010701	0,002780	0,000624
1%	0,078174	0,059038	0,029182	0,024914	0,012328	0,011020	0,002752	0,000684
5%	0,076497	0,044644	0,029586	0,026032	0,012720	0,011885	0,003863	0,002008
10%	0,083951	0,048873	0,034874	0,026253	0,013304	0,012112	0,005205	0,003731
20%	0,082267	0,065404	0,033007	0,027546	0,013453	0,015258	0,008265	0,007160

Tabelle 24: Ergebnisse des normierten euklidischen Abstandes bei Analyse der zweiten Ziffern

Die Ergebnisse der Analyse der ersten beiden Ziffern sind in Tabelle 25 dargestellt. Insgesamt ergeben sich auf Basis der Resultate ähnliche Schlussfolgerungen wie im Falle der ausschließlichen Betrachtung der ersten Ziffern: Hier fallen die Größen bis zu einem Stichprobenumfang von $n = 5.000$ Daten relativ groß aus. Dies hat auch hier zur Folge, dass eine echte Vergleichbarkeit lediglich im Falle $n \geq 5.000$ gegeben ist. Andernfalls nehmen die Werte zwar - mit Ausnahme des Datensatzes mit 500 Werten - stetig zu, jedoch liegen die Werte in einem divergierenden Bereich, sodass der Vergleich unterschiedlicher Stichprobengrößen zu falschen Ergebnissen führen könnte.

1. u. 2. Z.	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
0%	0,079180	0,055345	0,042794	0,034687	0,018490	0,013366	0,009927	0,009521
1%	0,082121	0,056368	0,042033	0,034219	0,018646	0,013633	0,009944	0,009585
5%	0,079630	0,052946	0,041764	0,035518	0,018736	0,015369	0,010803	0,010525
10%	0,078623	0,055128	0,041300	0,035364	0,019657	0,017086	0,012666	0,012645
20%	0,093445	0,067390	0,040952	0,036719	0,024568	0,021529	0,018586	0,018352

Tabelle 25: Ergebnisse des normierten euklidischen Abstandes bei Analyse der ersten beiden Ziffern

Insgesamt ist die Nutzung der normierten euklidischen Abstandsmessung als ein stichprobeninvariantes Instrument nur im Falle der ersten Ziffern sehr gut möglich. Hier sind lediglich 300 Daten erforderlich, um die Methode zuverlässig einzusetzen. Im Falle der Analyse der ersten beiden Ziffern sind hingegen mehr als 5.000 Daten notwendig. Werden diese Grenzen nicht eingehalten, oder sollen lediglich die zweiten Ziffern betrachtet werden, eignet sich die normierte euklidische Abstandsmessung nur, wenn Datensätze gleicher Stichprobengröße miteinander verglichen werden sollen.

4.3.7 Ergebnisse des Abstands über die arithmetischen Mittelwerte a^*

Bei der Analyse der Ergebnisse des Abstands über die arithmetischen Mittelwerte bzgl. der ersten Ziffern in Tabelle 26 fällt zunächst auf, dass die Ergebnisse des Datensatzes mit $n = 300$ Daten nicht plausibel erscheinen. Hier nimmt die Testgröße mit steigendem Anteil fraudulenter Daten ab. So liegt diese bei 20% manipulierter Daten bei 0,0233. Sind keine gleichverteilten Daten enthalten, liegt dieser Wert jedoch bei 0,0318. Ähnliche Beobachtungen lassen sich auch bei der Analyse der zweiten Ziffern (s. Tabelle 27) sowie der ersten beiden Ziffern (s. Tabelle

28) feststellen. Eine einfache Erklärung hierfür könnte darin liegen, dass im Falle des geringen Stichprobenumfanges von 300 Zahlen auf Grund des Gesetzes der Großen Zahlen eine relative Abweichung von der Benford-Verteilung beobachtet werden kann, wie sie auch zuvor bereits beschrieben wurde. Diese kann entsprechend dazu führen, dass der Mittelwert anfänglich stark von dem Mittelwert der Benford-verteilten Ziffern abweicht, was wiederum durch einen Überhang der kleinen Ziffern begründet sein könnte. Abbildung 9 in Abschnitt 4.3 bestätigt diese Vermutung. Durch das Hinzufügen gleichverteilter Daten wird dieser Überhang abgeschwächt, sodass sich auch der Mittelwert der Ziffern stärker an den Mittelwert der Benford-Verteilung annähert. Da die betrachtete Kennzahl die Abweichungen von der Benford-Verteilung auf Basis der arithmetischen Mittelwerte ermittelt, weisen die Ergebnisse entsprechende Verzerrungen auf. Für die Ergebnisinterpretation wird deshalb im Falle des Abstandes über die arithmetischen Mittelwerte auf den Einbezug der Resultate des Datensatzes mit $n = 300$ Daten verzichtet.

Ansonsten ergibt sich auch hier ein ähnliches Bild wie bei den vorherigen Abstandsmaßen: Der kleine Stichprobenumfang von $n = 150$ Zahlen weist insgesamt relativ große Werte auf. Auch wenn dieser Effekt im Vergleich zu der normierten euklidischen Abstandsmessung schwächer ausfällt, sollte auch hier darauf verzichtet werden, Ergebnisse dieses kleinen Datensatzes mit denen größerer Umfänge zu vergleichen. Für $n \geq 500$ Daten liegen die Ergebnisse hingegen in einem sehr ähnlichen Bereich. Wird diese Voraussetzung erfüllt, kann das Verfahren jedoch als stichprobeninvariant angesehen werden.

1. Zif.	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
0%	0,021541	(0,031819)	0,002835	0,000043	0,004503	0,006001	0,000051	0,000208
1%	0,027536	(0,028221)	0,000043	0,003555	0,001230	0,008519	0,000042	0,002991
5%	0,039527	(0,028821)	0,003555	0,024779	0,009994	0,021037	0,014213	0,014243
10%	0,053917	(0,003040)	0,021181	0,029275	0,023376	0,036901	0,027662	0,028438
20%	0,097084	(0,023340)	0,040966	0,043125	0,057370	0,062676	0,056596	0,056146

Tabelle 26: Ergebnisse des Abstands über die arithmetischen Mittelwerte bei Betrachtung der ersten Ziffern

Wie im Falle der vorherigen Abstandsmaße lässt sich auch hier bei der Analyse der zweiten Ziffern in Tabelle 27 feststellen, dass eine optimale Vergleichbarkeit und eine damit verbundene Stichprobeninvarianz erst bei Einbezug von ca. 1.000 Daten zu beobachten ist. Liegen

weniger als 500 Daten vor, so sind die Ergebnisgrößen allesamt relativ hoch, sodass von Interpretationen auf der Basis vergleichender Analysen stark unterschiedlicher Stichprobenumfänge abgesehen werden sollte. Sind die Stichprobengrößen hingegen nicht allzu weit voneinander entfernt, erscheint auch hier der Einsatz des Abstands über arithmetische Mittelwerte als durchaus geeignet.

2. Zif.	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
0%	0,086036	0,048333	0,018412	0,006107	0,000206	0,003285	0,001278	0,000060
1%	0,092962	0,054567	0,021737	0,004652	0,000829	0,003638	0,001301	0,000657
5%	0,094347	0,042793	0,021737	0,012386	0,006111	0,005488	0,005251	0,003495
10%	0,098503	0,047641	0,032957	0,002367	0,010973	0,008418	0,008216	0,006708
20%	0,124823	0,069112	0,046671	0,003821	0,013633	0,016417	0,014404	0,012924

Tabelle 27: Ergebnisse des Abstands über die arithmetischen Mittelwerte bei Betrachtung der zweiten Ziffern

Tabelle 28 präsentiert die Resultate bei Betrachtung der ersten beiden Ziffern. Aus den obigen Gründen wird auch hier vom Einbezug des Datensatzes mit 300 Zahlen abgesehen. Für die restlichen Ergebnisse ergeben sich im Wesentlichen die gleichen Schlussfolgerungen wie bei der ausschließlichen Betrachtung der ersten Ziffern: Ab 500 Daten liegen die Ergebnisse bei gleichen relativen Anteilen gleichverteilter Daten in einem ähnlichen Bereich, sodass man hier durchaus von einer Invarianz bzgl. des Stichprobenumfanges ausgehen kann. Aus diesem Grund ist das Instrumentarium auch prinzipiell für eine Analyse der ersten beiden Ziffern geeignet. Der Datensatz mit lediglich 150 Zahlen ist hingegen nur schlecht mit den größeren Datensätzen vergleichbar.

1. u. 2. Z.	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
0%	0,012971	(0,025433)	0,004076	0,000526	0,004161	0,005784	0,000149	0,000187
1%	0,017937	(0,021626)	0,001692	0,002901	0,001198	0,008130	0,000142	0,002805
5%	0,028862	(0,023116)	0,005003	0,023791	0,009684	0,019799	0,013499	0,013386
10%	0,041774	(0,000997)	0,022119	0,026754	0,022388	0,034632	0,026113	0,026707
20%	0,079406	(0,026986)	0,041421	0,039385	0,053886	0,058991	0,053235	0,052703

Tabelle 28: Ergebnisse des Abstands über die arithmetischen Mittelwerte bei Betrachtung der ersten beiden Ziffern

Insgesamt bleibt festzuhalten, dass auch der Abstandsmesser über die arithmetischen Mittelwerte bei Einhaltung einer gewissen Mindestgröße der Datensätze ein invariantes Verhalten hinsichtlich der Stichprobengröße aufweist. Demzufolge kann die Applikation im Falle einer Analyse der ersten Ziffern, der zweiten Ziffern sowie der ersten beiden Ziffern in Kombination empfohlen werden, sofern die genannte Bedingung erfüllt wird. Ist dies nicht der Fall spricht jedoch nichts gegen den Einsatz, sofern sich die Stichprobengrößen nicht allzu stark unterscheiden. Es sei jedoch auch darauf hingewiesen, dass unter bestimmten Umständen die attestierte Abweichung mit einem steigenden Manipulationsgrad abnimmt, da ein Überhang von kleinen bzw. großen Zahlen in Verbindung mit der Betrachtung des Mittelwertes als Basis der Berechnung durch Manipulationen kompensiert werden könnten. Diese Tatsache könnte z.B. dazu führen, dass die Größe a^* widersprüchliche Ergebnisse im Vergleich mit anderen Verfahren liefert.

4.3.8 Ergebnisse der Manhattan-Distanz

Von der Grundidee ist die Berechnung der Manhattan-Distanz derjenigen der euklidischen Abstandsmessung sehr ähnlich, was bereits in Abschnitt 3.4.8 erläutert wurde. Entsprechend ist es wenig verwunderlich, dass die Ergebnisse der ersten Ziffern – dargestellt in Tabelle 29 – sich nicht grundsätzlich von denjenigen der normierten euklidischen Abstandsmessung unterscheiden: Zunächst weist der Datensatz mit $n = 150$ Zahlen relativ große Werte auf, die selbst mit dem Datensatz mit $n = 300$ Zahlen nur schwer vergleichbar sind. Für alle anderen Datensätze ist ein Ergebnisvergleich lediglich für Stichproben ähnlicher Größenordnungen gegeben. So können die Werte der Datensätze $n = 300$ und $n = 1.000$ sehr plausibel miteinander verglichen werden, da die Ergebnisbereiche übereinstimmen. Möchte man jedoch den 500 Zahlen umfassenden Datensatz mit dem größten Datensatz mit 1 Mio. Zahlen vergleichen, so stößt man schnell auf Widersprüche: Die ideal Benford-verteilten Daten weisen bei dem kleinen Datensatz einen Wert von 0,0780 auf. Sind hingegen 10% der 1 Mio. Zahlen gleichverteilt, ergibt sich lediglich ein Wert von 0,0546. Demzufolge kann auch hier nur von einer eingeschränkten Stichprobeninvarianz gesprochen werden.

1. Zif.	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
0%	0,121577	0,070291	0,077997	0,042744	0,027578	0,025439	0,009465	0,001805
1%	0,130894	0,073156	0,077575	0,042681	0,033178	0,028524	0,009525	0,006276
5%	0,144227	0,086489	0,083635	0,070788	0,041672	0,048340	0,028163	0,027374
10%	0,172531	0,093156	0,103651	0,074788	0,059357	0,073740	0,051720	0,054560
20%	0,254729	0,124607	0,101741	0,098425	0,115140	0,121320	0,106600	0,107484

Tabelle 29: Ergebnisse der Manhattan-Distanz bei Analyse der ersten Ziffern

Tabelle 30 fasst die Ergebnisse hinsichtlich einer Betrachtung der zweiten Ziffern zusammen. Auch hier ergeben sich die gleichen Schlussfolgerungen wie im Falle der ersten Ziffern: Der kleinste Datensatz weist im Vergleich zu den restlichen Datensätzen deutlich größere Werte auf. Belässt man es insgesamt bei dem Vergleich von Stichproben einer ähnlichen Größenordnung, erscheint der Einsatz auch hier als durchweg empfehlenswert.

2. Zif.	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
0%	0,204048	0,140894	0,068933	0,077912	0,031655	0,026427	0,007334	0,001566
1%	0,204048	0,148202	0,075642	0,069912	0,030395	0,027427	0,007274	0,001629
5%	0,198086	0,106857	0,068933	0,067271	0,031655	0,031427	0,009076	0,005307
10%	0,217535	0,124135	0,083798	0,067912	0,033255	0,032427	0,012935	0,009771
20%	0,219188	0,157468	0,088190	0,069271	0,030680	0,041627	0,022155	0,018765

Tabelle 30: Ergebnisse der Manhattan-Distanz bei Analyse der zweiten Ziffern

Wirft man hingegen einen Blick auf die Ergebnisse bei Einbezug der ersten beiden Ziffern in Tabelle 31, ergeben sich differenzierte Beobachtungen: Die Resultate der sich unterscheidenden Stichprobengrößen liegen fast durchweg in divergierenden Wertebereichen. Diese nehmen mit einem Anstieg der Größe des Umfanges tendenziell ab. Dementsprechend kann hier nicht von einer Stichprobeninvarianz ausgegangen werden. Der Einsatz des Instrumentes ist lediglich für nahezu exakt gleiche Stichprobengrößen zu empfehlen.

1. u. 2. Z.	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
0%	0,592443	0,417377	0,297075	0,256638	0,124246	0,095609	0,052446	0,050759
1%	0,606777	0,424043	0,286536	0,252810	0,123940	0,096596	0,052437	0,049741
5%	0,590025	0,400112	0,280350	0,261677	0,117346	0,107323	0,054898	0,051984
10%	0,582889	0,407702	0,288375	0,260005	0,117336	0,114445	0,067319	0,066252
20%	0,677342	0,502916	0,284066	0,264713	0,150259	0,144383	0,110755	0,108627

Tabelle 31: Ergebnisse der Manhattan-Distanz bei Analyse der ersten beiden Ziffern

Im Allgemeinen lässt sich festhalten, dass die Abstandsmessung mittels der Manhattan-Distanz bei der Analyse der ersten sowie der zweiten Ziffern eine eingeschränkte Invarianz bzgl. des Stichprobenumfanges aufweist. Demzufolge sind in der Größenordnung ähnliche Datensätze optimal vergleichbar. Nimmt die Divergenz zu, so sinkt jedoch im Gegenzug die Vergleichbarkeit der Ergebnisse. Untersuchungen von allzu kleinen Stichprobenumfängen von weniger als 300 Daten sollten generell vermieden werden.

Sollen die ersten beiden Ziffern in Kombination analysiert werden, so weist die Manhattan-Distanz keine Invarianz bzgl. des Stichprobenumfanges auf. Der Einsatz sollte hier deshalb nur erfolgen, wenn die zu untersuchenden Datensätze im Umfang exakt gleich sind.

4.3.9 Ergebnisse der Tschebyscheff-Distanz

Die Ergebnisse der Tschebyscheff-Distanz der ersten Ziffern sind in Tabelle 32 dargestellt. Auffällig ist zunächst, dass sich die Ergebniswerte innerhalb eines Datensatzes bei unterschiedlichen Anteilen gleichverteilter Daten nicht zwingend unterscheiden. So weist der Datensatz mit 500 Zahlen im Falle 0%, 1% und 5% enthaltener manipulierter Daten den Wert 0,1019909 auf. Zu begründen ist dies damit, dass hier der maximale Abstand zwischen den empirischen und den hypothetischen Klassen als Kennzahl verwendet wird. Ändern sich jedoch die Häufigkeiten der Klasse mit maximalen Abstand im Zuge von Manipulationen nicht, so würde die Kennzahl von den Manipulationen unberührt bleiben, solange die Distanzen anderer Klassen nicht maximal würden. Abstrahiert von diesen einzelnen Ausnahmen, steigt die Testgröße jedoch im Allgemeinen mit zunehmendem Anteil fraudulenter Daten ebenfalls an, was grundsätzlich für die Anwendbarkeit der Kenngröße im Rahmen des Benford-Testes spricht.

Die Ergebnisse des Datensatzes des kleinsten Stichprobenumfanges weisen auch hier relativ große Werte auf, die eine Vergleichbarkeit mit den anderen Datensätzen unmöglich machen. Hinsichtlich der Invarianz des Stichprobenumfanges lässt sich beobachten, dass diese eingeschränkt unter der Voraussetzung vorliegt, dass die zu vergleichenden Stichprobenumfänge in einer ähnlichen Größenordnung liegen und sich lediglich um den Faktor 10 unterscheiden. Andernfalls könnten auch hier falsche Ergebnisse geschlussfolgert werden.

1. Zif.	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
0%	0,039720	0,022303	0,019909	0,013090	0,011909	0,007653	0,001411	0,000387
1%	0,046387	0,025637	0,019909	0,014090	0,011509	0,007753	0,001411	0,001697
5%	0,046387	0,018970	0,019909	0,017091	0,013630	0,011930	0,008640	0,009211
10%	0,042758	0,019425	0,021030	0,018030	0,023630	0,022230	0,017850	0,018758
20%	0,081030	0,026091	0,033030	0,030030	0,044430	0,037030	0,037630	0,037473

Tabelle 32: Ergebnisse der Tschebyscheff-Distanz bei Analyse der ersten Ziffern

Die Ergebnisse der Analyse der zweiten Ziffern zeigt Tabelle 33. Hier fällt auf, dass sich die Wertebereiche der Testgröße hinsichtlich der Stichprobenumfänge durchaus unterscheiden. Demzufolge ist ein Ergebnisvergleich nicht möglich, wenn die Datensätze in deren Größe divergieren. Dieses Distanzmaß sollte folglich zur Analyse der zweiten Ziffern nur eingesetzt werden, wenn die Eigenschaft der Stichprobeninvarianz nicht erforderlich ist.

2. Zif.	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
0%	0,040321	0,029097	0,017679	0,012821	0,006523	0,006623	0,001210	0,000324
1%	0,040321	0,032430	0,017679	0,014821	0,006197	0,006823	0,001239	0,000448
5%	0,040321	0,029097	0,015692	0,014821	0,007679	0,007523	0,001989	0,001126
10%	0,046110	0,029097	0,019692	0,014670	0,008479	0,007523	0,002619	0,001962
20%	0,039443	0,042981	0,017679	0,019670	0,008430	0,008023	0,004139	0,003814

Tabelle 33: Ergebnisse der Tschebyscheff-Distanz bei Analyse der zweiten Ziffern

Werden die ersten beiden Ziffern in Kombination analysiert, so lassen sich hier ähnliche Auffälligkeiten beobachten, wie bei ausschließlicher Betrachtung der ersten Ziffern. Die Resultate der Berechnungen sind in Tabelle 34 dargestellt. Zwar weisen die kleinen Datensätze mit $n \leq$

500 Daten jeweils analog zu den Ergebnissen der zweiten Ziffern einen eigenen Wertebereich auf, jedoch liegt im Gegensatz dazu bei den größeren Stichprobenumfängen eine sehr gute Komparabilität der ähnlich großen Stichprobenumfänge vor, die auf eine vorhandene Invarianz hinsichtlich der Stichprobengröße schließen lässt.

1. u. 2. Z.	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
0%	0,026271	0,018059	0,015797	0,010271	0,008593	0,004693	0,005543	0,005576
1%	0,026802	0,018059	0,015797	0,010271	0,008993	0,004693	0,005573	0,005801
5%	0,026802	0,018059	0,016037	0,009189	0,009193	0,005793	0,006743	0,006808
10%	0,026271	0,018059	0,014037	0,010029	0,010193	0,006793	0,007643	0,008022
20%	0,028183	0,018841	0,015393	0,011789	0,012793	0,008893	0,009893	0,010401

Tabelle 34: Ergebnisse der Tschebyscheff-Distanz bei Analyse der ersten beiden Ziffern

Zusammenfassend eignet sich die Tschebyscheff-Distanz bei der Analyse der ersten sowie der ersten beiden Ziffern, um Berücksichtigung im Rahmen des Benford-Testes zu finden. Dies gilt vor allem deshalb, da Datensätze unterschiedlichen Stichprobenumfanges gut verglichen werden können, solange der Unterschied nicht deutlich ausgeprägt ist. Im Falle der zweiten Ziffern ist eine Eignung nur empfehlenswert, wenn Stichproben ähnlicher Größen miteinander verglichen werden sollen.

4.3.10 Ergebnisse des Verzerrungsfaktor-Modells

Die Untersuchungsergebnisse des von *Nigrini* entwickelten Verzerrungsfaktor-Modells sind in Tabelle 35 dargestellt. Die Berechnungen erfolgen hierbei ausschließlich auf Basis des Mittelwertes der beobachteten Zahlen. Aus diesem Grund werden die Ziffern nicht einzeln betrachtet, sondern vielmehr die generierten Zahlen als solches. Hierbei entspricht ein Ergebniswert jeweils der relativen Abweichung des Mittelwertes der empirischen Daten von dem Mittelwert, welcher bei einer idealen Benford-Verteilung hypothetisch auftritt. Ein Wert von 0,037 sagt demnach aus, dass der empirische Mittelwert 3,7% oberhalb des Mittelwertes ideal Benford-verteilter Zahlen liegt. Eine Größe von -0,105 signalisiert eine Unterschreitung dieses Wertes um 10,5%.

Zu erkennen ist, dass bei den kleinen Datensätzen mit $n \leq 300$ Zahlen selbst bei den idealen Benford-Daten Deviationen von 2,8% bzw. -3,5% auftreten. Die Ursache ist auch hier dem Gesetz der Großen Zahlen geschuldet, da die relativen Abweichungen mit Zunahme der Stichprobengröße abnehmen. Betrachtet man ausschließlich den Datensatz mit 300 Werten, so fällt ein weiterer Nachteil des Verfahrens auf: Die Ergebniswerte drehen mit zunehmendem Anteil manipulierter Daten von dem negativen in den positiven Bereich. Eine ähnliche Beobachtung konnte bereits bei der Größe a^* gemacht werden. Dies lässt darauf schließen, dass anfänglich zu viele kleine Zahlen in dem Datensatz waren, diese Auffälligkeit jedoch mit Zunahme des Anteils gleichverteilter Zahlen kompensiert wird. So könnte man fälschlicherweise zu dem Schluss kommen, dass der Datensatz mit 10% fraudulent Daten mit 0,055% Abweichung eine größere Annäherung an die Benford-Verteilung aufweist als der Datensatz ohne jegliche Manipulationen. Abgesehen davon liegt für die Datensätze mit $n \geq 500$ Zahlen eine relativ gute Vergleichbarkeit der Ergebnisse vor, die auf eine Stichprobeninvarianz schließen lassen. Eine Ausnahme bildet lediglich der größte Datensatz mit 1 Mio. Zahlen: Hier liegen die Ergebniswerte bei vier von fünf Datensätzen bei 0,00016, was darauf schließen lässt, dass Manipulationen in sehr großen Datensätzen nur unzureichend erkannt werden.

	n = 150	n = 300	n = 500	n = 1.000	n = 5.000	n = 10.000	n = 100.000	n = 1.000.000
0%	0,02804	-0,03528	0,00802	0,00018	-0,00628	0,00900	0,00007	0,00016
1%	0,03586	-0,02948	0,00437	0,00550	-0,00171	0,01262	0,00006	0,00420
5%	0,05284	-0,03192	0,00965	0,03790	0,01509	0,03064	0,02071	0,00016
10%	0,07316	0,00554	0,03617	0,04262	0,03477	0,05356	0,04020	0,00016
20%	0,13171	0,04580	0,06593	0,06226	0,08345	0,09121	0,08214	0,00016

Tabelle 35: Ergebnisse des Verzerrungsfaktor-Modells

4.4 Auswertung und Schlussfolgerungen der Untersuchung

In Tabelle 36 ist zunächst die generelle Eignung der verschiedenen Testgrößen und Maßzahlen zur Anwendung im Rahmen des Benford-Testes dargestellt. Zu erkennen ist, dass bei dem Großteil der Verfahren nichts gegen den Einsatz zur Manipulationsaufdeckung spricht. Lediglich bei der Analyse der ersten beiden Ziffern sollte auf die Anwendung des Chi²-Anpassungstestes sowie des Kuiper-Testes verzichtet werden, da die erhaltenen Resultate schnell zu falschen Interpretationen führen können.

Die Z-Statistik weist wie in Abschnitt 4.3.4 dargelegt lediglich eine eingeschränkte Eignung auf. Zum einen ist es weder ratsam, bei der Überschreitung des kritischen Wertes von nur einer Klasse die Benford-Verteilung zu verwerfen, zum anderen ist auch die Interpretation der maximal auftretenden Testgröße nicht empfehlenswert, da sich diese stets nur auf eine Merkmalsklasse bezieht und somit auch hier falsche Schlüsse gezogen werden könnten. Lediglich der Vergleich der Anzahl an Klassen mit signifikanten Abweichungen könnte ein sinnvolles Vergleichskriterium sein, jedoch lässt sich auch hier nicht eindeutig feststellen, wann eine Benford-Verteilung noch unterstellt werden kann und ab wann diese zu verwerfen wäre. Aus diesem Grund sollte die Z-Statistik eher dazu eingesetzt werden, im Falle erkannter Abweichungen diejenigen Ziffern zu identifizieren, die die größten Abweichungen von der hypothetischen Klassenwahrscheinlichkeit aufweisen. Dies kann sehr nützlich sein, um z.B. Manipulationsmuster wie gezielte Auf- oder Abrundungen zu identifizieren und zu belegen.

Auch weist das Verzerrungsfaktor-Modell des amerikanischen Forschers *Nigrini* lediglich eine eingeschränkte Eignung zur Manipulationsaufdeckung auf: Zwar ermittelt die Größe auf Basis des Abstands der Mittelwerte den Grad der Abweichung, sodass eine höhere Deviation in der Regel auch mit betragsmäßig höheren Testwerten einhergeht. Es können sogar Aussagen darüber getroffen werden, ob zu viele kleine oder zu viele große Zahlen in dem Datensatz vorhanden sind. Da jedoch kleine Datensätze auch ohne Manipulationen aufgrund des Gesetzes der Großen Zahlen kleinere Abweichungen von der idealen Benford-Verteilung aufweisen, kann es wie oben gezeigt vorkommen, dass beispielsweise die Anzahl kleiner Zahlen in dem Datensatz überwiegt. Würden nun Zahlen so manipuliert, dass die Anzahl großer Zahlen zunimmt, würde entgegen der Sinnhaftigkeit der Wert der Testgröße abnehmen. Demzufolge sind die Ergebnisse des Verzerrungsfaktor-Modells stets mit Vorsicht zu genießen.

	1. Ziffern	2. Ziffern	1. u. 2. Ziffern
Chi²-Anpassungstest	✓	✓	✗
Kolmogorow-Smirnow-Test	✓	✓	✓
Kuiper-Test	✓	✓	✗
Z-Statistik	(✓)	(✓)	(✓)
Normierter Chi²-Wert	✓	✓	✓
Normierter euklidischer Abstand	✓	✓	✓
Abstand über die arithmetischen Mittelwerte	✓	✓	✓
Manhattan-Distanz	✓	✓	✓
Tschebyscheff-Distanz	✓	✓	✓
Verzerrungsfaktor-Modell	(✓)	(✓)	(✓)

Tabelle 36: Generelle Eignung der Testgrößen zur Applikation im Rahmen des Benford-Testes

Des Weiteren stellt Tabelle 37 dar, ob die einzelnen Testverfahren auch in der Praxis eine Invarianz hinsichtlich des Stichprobenumfanges aufweisen. Wie es zu erwarten war, weisen hierbei sowohl der Chi²-Anpassungstest als auch der Kuiper-Test sehr hohe Abhängigkeiten der Testgröße von der Stichprobengröße auf. V.a. bei großen Stichprobenumfängen genügen hierbei schon kleinste Abweichungen, um die Testgröße exorbitant ansteigen zu lassen und damit auch die Nullhypothese signifikant abzulehnen. Möchte man hierbei die Werte zwischen verschiedenen Datensätzen vergleichen, so sollten sich diese in ihrer Größe nicht unterscheiden. Andernfalls ist die Komparabilität der Werte nicht sichergestellt. Auch die Werte der Z-Statistik sind stark abhängig von der Stichprobengröße. Da die Eignung im Rahmen des Benford-Testes jedoch ohnehin nur eingeschränkt möglich ist, sollte der Einsatz nur dazu dienen, die Ziffern mit den größten Abweichungen zu identifizieren.

Eine echte Stichprobeninvarianz weisen zum einen der KS-Test sowie der Abstand über die arithmetischen Mittelwerte auf. Hierbei sind die Ergebnisse auch vergleichbar, wenn sich die Umfänge der Datensätze stark unterscheiden.

Bei den restlichen Größen liegt eine eingeschränkte Stichprobeninvarianz vor. Dies bedeutet konkret, dass eine Unabhängigkeit vom Stichprobenumfang vorliegt, sofern die Unterschiede

nicht allzu groß sind. Als Richtwert kann man davon ausgehen, dass eine sehr gute Vergleichbarkeit der Werte zu beobachten ist, sofern sich die Umfänge maximal um den Faktor zehn unterscheiden. Sind die Abstände größer, sollten Interpretationen auf Basis des Vergleichs mehrerer Werte lediglich mit Vorsicht genossen werden. Als Ausnahme sei an dieser Stelle noch auf den normierten euklidischen Abstand sowie der Tschebyscheff-Distanz verwiesen. Auf deren Anwendung auf die zweiten Ziffern sollte nur zurückgegriffen werden, wenn die Stichprobenumfänge übereinstimmen, da die Ergebniswerte mit steigendem Manipulationsgrad nur sehr geringfügig ansteigen.

	Stichprobeninvarianz
Chi²-Anpassungstest	✗
Kolmogorow-Smirnow-Test	✓
Kuiper-Test	✗
Z-Statistik	✗
Normierter Chi²-Wert	(✓)
Normierter euklidischer Abstand	(✓)
Abstand über die arithmetischen Mittelwerte	✓
Manhattan-Distanz	(✓)
Tschebyscheff-Distanz	(✓)
Verzerrungsfaktor-Modell	(✓)

Tabelle 37: Praktische Erfüllung der Eigenschaft der Stichprobeninvarianz einzelner Testverfahren

Die Anforderungen an die Größen der Stichprobenumfänge der einzelnen Verfahren zeigt Tabelle 38. Die Ergebnisse stellen hierbei Beobachtungswerte aus dem vorherigen Abschnitt dar und sollten lediglich als ungefähre Richtwerte aufgefasst werden. Lediglich im Falle des Chi²-Anpassungstestes, bei welchem die Größen mit einem Sternchen markiert wurden, sind die Werte angeführt, die der Theorie nach mindestens vorhanden sein müssen, um korrekte Ergebnisse zu erhalten. Die Eintragung (-) weist jeweils darauf hin, dass in dieser Fallkonstellation eine Anwendung des Instrumentariums nicht empfohlen werden kann. Im Allgemeinen ist es bei der Analyse der ersten Ziffern empfehlenswert einen minimalen Stichprobenumfang von 300 bis 500 Zahlen vorliegen zu haben, sodass der Großteil der Verfahren Anwendung finden

kann. Sollen die zweiten Ziffern analysiert werden, ist die Anzahl von ca. 300 bis 1.000 Werten empfehlenswert. Zur Untersuchung der ersten beiden Ziffern in Kombination sind nur sechs der Verfahren tendenziell zum Einsatz geeignet. Die Mindeststichprobenumfänge sind hierbei recht unterschiedlich. Sie reichen von mindestens 300 bis zu der Anforderung von 5.000 Zahlen im Falle des normierten euklidischen Abstandes.

	1.Ziffern	2. Ziffern	1. u. 2. Ziffern
Chi²-Anpassungstest	≥ 110*	≥ 59*	-
Kolmogorow-Smirnow-Test	≥ 300	≥ 300	≥ 300
Kuiper-Test	keine	keine	-
Z-Statistik	≥ 5.000	≥ 100.000	-
Normierter Chi²-Wert	≥ 300	≥ 500	≥ 500
Normierter euklidischer Abstand	≥ 300	keine	≥ 5.000
Abstand über die arithmetischen Mittelwerte	≥ 500	≥ 1.000	≥ 500
Manhattan-Distanz	≥ 300	≥ 500	keine
Tschebyscheff-Distanz	≥ 300	keine	≥ 1.000
Verzerrungsfaktor-Modell	$500 \leq n \leq 100.000$		

Tabelle 38: Auf Basis der Untersuchung empfohlene Größe des zu untersuchenden Stichprobenumfanges der einzelnen Testgrößen

Insgesamt bleibt festzuhalten, dass v.a. im Fall der statistischen Anpassungstests stets das Gesetz der Großen Zahlen Berücksichtigung findet. Dies führt dazu, dass zur Ablehnung der Nullhypothese bei kleinen Stichprobenumfängen eine außerordentlich große Anzahl an Manipulationen vorliegen muss. Bei großen Stichprobenumfängen werden hingegen bereits kleinste Anteile an manipulierten Daten erkannt und die Benford-Verteilung verworfen. Bei den Abstandsmaßen fällt dieser Punkt hingegen kaum ins Gewicht. Aus diesem Grund erscheint es tendenziell sinnvoll, auf die Interpretationen statistisch signifikanter Entscheidungen im Rahmen des Benford-Testes nur eingeschränkt Wert zu legen. Da alle der vorgestellten Verfahren grundsätzlich unter den genannten Voraussetzungen geeignet sind, um im Rahmen eines Vergleiches von zwei oder mehr Stichproben zur Anwendung zu kommen, erscheint es ratsamer, eine Kombination sämtlicher der oben vorgestellten Testverfahren und Maßzahlen zur Auswertung zu verwenden. Weisen mehrere Größen in die gleiche Richtung, so steigt auch die Robustheit der

Ergebnisse systematisch an. Gibt es hingegen Widersprüche zwischen den einzelnen Verfahren, so sollten die Resultate als weniger schlagkräftig angesehen werden. Lediglich die Z-Statistik und das Verzerrungsfaktormodell sollten aufgrund ihrer eingeschränkten Verwendbarkeit lediglich in Ergänzung zu den restlichen Verfahren zur Anwendung kommen.

Bei der Analyse wird zudem empfohlen, wenn möglich sowohl die ersten als auch die zweiten Ziffern zu untersuchen. Somit können die Ergebnisse zusätzlich ein weiteres Mal validiert werden. Die Betrachtung der ersten beiden Ziffern in Kombination bringt hingegen kaum Vorteile, da Manipulationen hier im Allgemeinen unzuverlässiger entdeckt werden. Aus diesem Grund wird hierauf auch im Rahmen der nachfolgenden empirischen Untersuchungen verzichtet.

5. Vergleich der Qualität historischer makroökonomischer Daten der Bundesrepublik Deutschland und ihrer Vorgängerstaaten

5.1 Grundlagen des Datenmaterials

Der gesamte Datenbestand, der im Rahmen dieser empirischen Untersuchung verwendet wird, wurde aus den statistischen Jahrbüchern Deutschlands der Jahre 1918 bis einschließlich 1990 erhoben. Diese Jahrbücher erscheinen grundsätzlich bis auf wenige Ausnahmen jährlich und beinhalten neben Daten der Volkswirtschaftlichen Gesamtrechnung u.a. Wirtschaftsdaten und Daten der Sozialstatistik. Sie sind öffentlich zugänglich und frei abrufbar.¹⁹⁶ Erhoben wurden hierbei ausschließlich Daten, die mit der wirtschaftlichen Leistungsfähigkeit des Staates in Verbindung stehen. Insbesondere wurde auf den Einbezug sämtlicher Daten verzichtet, die in Verbindung mit der Sozialstatistik ohne wirtschaftlichen Kontext stehen. Konkret wurden für die Datenerhebung innerhalb der einzelnen Jahre die statistischen Jahrbücher des Deutschen Reiches (1918 - 1942), der Deutschen Demokratischen Republik (1955 bis 1990) sowie der Bundesrepublik Deutschland (1952 - 1990) herangezogen.

Im Rahmen der Datenerhebung wurden alle verfügbaren Daten verwendet, die in einer Währungseinheit vorlagen und somit - durch den Geldwert gekennzeichnet - unmittelbar die Höhe der jeweiligen wirtschaftlichen Leistung angeben. Auf weitere zufallsbasierte Selektionen der Daten wurde verzichtet.

Da sämtliche betrachtete Daten in unmittelbarem Zusammenhang mit der wirtschaftlichen Leistungsfähigkeit der jeweiligen Regierungs- und Staatsform stehen, besteht bei diesen grundsätzlich die Gefahr, Manipulationen und Falschangaben zu beinhalten: Schließlich ist jede Regierung jederzeit bestrebt, die Leistungsfähigkeit des Staates und damit auch die eigene Leistungsfähigkeit stets so positiv wie möglich nach außen zu propagieren. Einzelne Regierungen wie z.B. das NS-Regime in der Zeit von 1933 bis 1945 könnten zudem die Motivation besitzen, bestimmte wirtschaftliche oder staatliche Geschehnisse - z.B. die systematische Vorbereitung auf einen bevorstehenden Krieg - in den veröffentlichten Daten gezielt zu verschleiern oder deren ökonomische Konsequenzen zu beschönigen. Aus diesem Grund eignen sich die untersuchten Daten im besonderen Maße, um im Rahmen des Benford-Testes genauer untersucht zu werden und mittels dessen Methodik deren Qualität zu überprüfen. Der Benford-Test müsste

¹⁹⁶ Die statistischen Jahrbücher finden sich in digitaler Form unter <http://www.digizeitschriften.de/dms/toc/?PID=PPN514401303> (Für diese Untersuchung abgerufen in dem Zeitraum: 01.06. - 01.10.2015).

hierbei grundsätzlich die Eignung besitzen, solche gezielten Eingriffe in die statistischen Daten zu erkennen und aufzudecken. Sofern dies der Fall ist, wäre dies ein weiterer Beweis für die Eignung des Benford-Testes zur Manipulationsaufdeckung in volkswirtschaftlichen Daten.

5.2 Deskriptive Statistik und Untersuchungsdesign

Wie bereits dargelegt, wurden für diese empirische Untersuchung lediglich Daten verwendet, die in Zusammenhang mit der wirtschaftlichen Leistungsfähigkeit der Staaten stehen, da die Wahrscheinlichkeit für vorhandene Manipulationen hier am größten ist. Aus diesem Grund wurden nur bestimmte Kapitel der statistischen Jahrbücher in die Datenerhebung einbezogen. Namentlich waren dies für das Deutsche Reich von 1918 bis 1945 (Weimarer Republik und NS-Regime) die folgenden Abschnitte:

1. Land- und Forstwirtschaft und Fischerei (ab 1934: Land und Forstwirtschaft)
2. Viehwirtschaft und Veterinärwesen (ab 1934: Viehwirtschaft, Fischerei, Veterinärwesen)
3. Gewerbe
4. Verkehr
5. Auswärtiger Handel (ab 1939: Außenhandel)
6. Geld- und Kreditwesen
7. Finanzwesen (ab 1934: Öffentliche Finanzwirtschaft)
8. Versicherungswesen
9. Berufsverbände (ab 1921: Berufsvertretungen)
10. Volkswirtschaftliche Bilanzen (ab 1933)

Für die Deutsche Demokratische Republik (1955 bis 1990) wurden die Kapitel

1. Volkswirtschaftliche Bilanz (1958: Gesellschaftliches Gesamtprodukt und Nationaleinkommen; 1959 - 1976: Volkswirtschaftliche Übersichten; ab 1977: Volkswirtschaftliche Kennziffern)
2. Industrie ohne Bau (ab 1956: Industrie)
3. Bauindustrie (ab 1956: Bauwirtschaft)
4. Handwerk und Kleinindustrie (ab 1956: Handwerk)
5. Land- und Forstwirtschaft (1969 bis 1972: Landwirtschaft und Nahrungsgüterwirtschaft)

6. Verkehr und Nachrichtenwesen (1956 - 1959: Verkehr; 1960 – 1964: Transport- und Nachrichtenwesen; ab 1965: Verkehr, Post und Fernmeldewesen)
7. Binnenhandel (ab 1956: Handel)
8. Außenhandel und innerdeutscher Handel (ab 1963: Außenhandel; 1969 - 1972: Außenwirtschaft)
9. Finanzen (ab 1966: Finanzen, Preise, Verbrauch)

in der Datenerhebung berücksichtigt. Für die Bundesrepublik Deutschland (1952 – 1990) gliederten sich die betrachteten Kapitel wie folgt:

1. Land- und Forstwirtschaft, Fischerei
2. Industrie und Handwerk (ab 1976: Produzierendes Gewerbe)
3. Bauwirtschaft, Gebäude, Wohnungen (1958 - 1975: Bauwirtschaft, Bautätigkeit und Wohnungen; ab 1976: Bautätigkeit und Wohnungen)
4. Binnenhandel (1953 - 1960: Binnenhandel und Interzonenhandel; 1961 - 1976: Groß- und Einzelhandel, Gastgewerbe, Fremdenverkehr; ab 1977: Handel, Gastgewerbe, Reiseverkehr)
5. Außenhandel
6. Verkehr
7. Geld und Kredit (ab 1977: zusammengefasst unter „Geld, Kredit, Versicherungen“)
8. Versicherungen (ab 1977: zusammengefasst unter „Geld, Kredit, Versicherungen“)
9. Öffentliche Sozialleistungen (ab 1973: Sozialleistungen)
10. Öffentliche Finanzen (ab 1962: Finanzen und Steuern)
11. Volkswirtschaftliche Gesamtrechnungen
12. Zahlungsbilanz

Gut zu erkennen ist hierbei, dass der Aufbau der statistischen Jahrbücher und damit auch der Kern der enthaltenen Zahlen zwischen den Staatsformen eine hohe Übereinstimmung aufweisen und eine Vergleichbarkeit der Datensätze inhaltlich somit sehr gut möglich ist. Bei Erhebung der Zahlen war im Kontext der Anwendung des Benford-Testes jedoch insbesondere darauf zu achten, dass die einbezogenen Daten sich nicht durch andere Zahlen des Datensatzes mittels mathematischer Operationen wie z.B. die Addition oder Subtraktion reproduzieren lassen. Dies würde andernfalls zu einer Doppelerfassung etwaiger Abweichungen von der Benford-Verteilung führen: Setzt sich das Bruttoinlandsprodukt als Beispiel in der Entstehungsrechnung aus dem Produktionswert abzüglich Vorleistungen plus Nettogütersteuern zusammen

und liegen alle vier Werte vor, so ist es lediglich zulässig die ersten drei Größen im Rahmen einer Datenerhebung mit einzubeziehen. Denn dem Anschein nach würden potentielle Manipulationen mit sehr großer Wahrscheinlichkeit auf der untersten Aggregationsebene (hier z.B. beim Produktionswert) durchgeführt. Eine mögliche Deviation von der Benford-Verteilung würde jedoch dann auch unmittelbar auf der höchsten Aggregationsebene (hier bei der Größe Bruttoinlandsprodukt) festzustellen sein und würde somit doppelt erfasst. Gleiches gilt für den Fall, dass eine Größe sich auf unterschiedliche Weisen berechnen lässt. Um das obige Beispiel fortzuführen, existiert so für das BIP zusätzlich zu der Entstehungsrechnung auch eine Verwendungsrechnung, mittels welcher diese Größe auch in Konsumausgaben zzgl. Bruttoinvestitionen zzgl. Außenbeitrag aufgeschlüsselt werden kann. Demnach wirkt sich folglich eine Manipulation in Posten der Entstehungsrechnung auch auf die Verwendungsrechnung aus, sodass die Manipulationen ebenfalls doppelt erfasst würden. Somit ist zwingend darauf zu achten, solche Doppelerhebungen zu vermeiden und so den Umfang der einbezogenen Daten zu reduzieren. Im Falle der hier vorliegenden Daten trat häufig genau dieses Problem auf: So wurde beispielsweise die Produktionsleistung in der Industrie auf der einen Seite nach einzelnen Branchen und andererseits nach Bundesländern aufgeschlüsselt. Erhoben wurden dann lediglich die Zahlen der einzelnen Branchen, nicht jedoch die Gesamtsummen der Produktionsleistungen sowie die Werte der einzelnen Bundesländer oder Regionen.

Grundsätzlich müssen sämtliche in Abschnitt 3.2 genannten Voraussetzungen an den Datensatz erfüllt werden, damit eine Anwendung des Benford-Testes überhaupt erfolgreich möglich ist. In dieser Untersuchung ist dieser Aspekt jedoch weitestgehend unproblematisch, da der Datensatz weder feste Ober- und Untergrenzen enthält noch auf Basis nicht-zufallsbasierter Erhebungen entstanden ist. Da alle Zahlen entweder in der Einheit Reichsmark, Rentenmark, Ostmark oder Deutsche Mark vorliegen, ist ebenfalls die Kongruenz der Einheitsdimensionen innerhalb aller erhobenen Zahlen sichergestellt. Statistische Kennwerte befinden sich nicht innerhalb des Datensatzes.

Eine Übersicht über die statistischen Kennzahlen inklusive des Stichprobenumfanges des Gesamtdatensatzes sowie der einzelnen Staatsformen gibt Tabelle 39. Gut zu erkennen ist, dass der Gesamtstichprobenumfang deutlich über den in Abschnitt 4.3 empfohlenen 300 Werten liegt, sodass diese Voraussetzung ebenfalls erfüllt ist. Auch liegen sämtliche Stichprobenumfänge in einem ähnlichen Bereich und unterscheiden sich maximal um den Faktor fünf. Dem-

nach ist die Anwendung aller in Abschnitt 4.3. vorgestellten Verfahren im Rahmen eines relativen Ergebnisvergleichs uneingeschränkt möglich. Der Median ist in allen Fällen kleiner als der Mittelwert und die Schiefe deutlich größer null, sodass hier ebenfalls nichts gegen die Anwendung des Benford-Testes spricht.

Staatsform	Stichproben- umfang n	Mittelwert	Median	Schiefe
<i>Weimarer Republik</i>	53.189	30.300.626,41	3.137,58	23,99
<i>NS-Regime</i>	20.523	59.240,25	743,74	13,60
<i>Deutsche Demokratische Republik</i>	45.597	18.203.363,27	8.492,42	16,64
<i>Bundesrepublik Deutschland</i>	96.769	155.511,97	1.538,95	21,13
Gesamt	199.978	8.083.860	3.478	18,84

Tabelle 39: Stichprobenumfang und statistische Kennzahlen des Gesamtdatensatzes und der einzelnen Staatsformen

Anhang II (Tabelle 75 bis 78) zeigt die statistischen Kennzahlen sowie die Stichprobenumfänge der Einzeljahre in Abhängigkeit von der jeweiligen Staatsform. Auch hier sind stets mehr als 300 Zahlen verfügbar. Vergleicht man die Stichprobenumfänge der Einzeljahre liegen diese zum größten Teil zwischen ca. 1.000 und 5.000 Zahlen und bieten demnach auch eine hervorragende Basis für die Anwendung sämtlicher Testverfahren. Selbst bei den eingeschränkt stichprobeninvarianten Größen, ist somit eine exzellente Vergleichbarkeit zwischen den Einzeljahren gegeben. Ebenfalls sprechen auch hier die statistischen Kennzahlen nicht gegen eine Benford-Verteilung: Die Verteilungen sind allesamt rechtsschief, da die Schiefe größer als null ist. Des Weiteren ist der Median in sämtlichen Fällen kleiner als der Mittelwert. Aus diesem Grund kann der Benford-Test nicht nur auf die Gesamtdatensätze der einzelnen Staatsformen angewendet werden, sondern ebenfalls eine genaue Analyse der Einzeljahre vorgenommen werden.

Wie in Abschnitt 4.4 dargelegt wurde, ist es grundsätzlich empfehlenswert, den Benford-Test auf die ersten und die zweiten Ziffern anzuwenden, sofern dies möglich ist. Dies ist der Fall, wenn zu erwarten ist, dass der zu analysierende Datensatz im nicht-manipulierten Fall grundsätzlich auch Benford-verteilt ist. Demnach wird in Abschnitt 5.3.1 zunächst überprüft, ob der aggregierte Datensatz die Benford-Verteilung für die ersten und zweiten Ziffern aufweist. Angewendet werden im Rahmen der Untersuchung, wie vorgeschlagen, sämtliche in Abschnitt 3.3 vorgestellten Verfahren. Die Ergebnisse der Z-Statistik werden eignungsbedingt lediglich dazu

verwendet, diejenige Ziffer mit der größten Benford-Abweichung zu ermitteln. Das Verzerrungsfaktormodell kommt lediglich ergänzend zur Anwendung.

Auch wenn die Interpretation der statistisch signifikanten Entscheidung wie in Abschnitt 4.4 dargelegt, nicht sinnvoll erscheint und es ratsamer ist, sämtliche Verfahren gleichgewichtet heranzuziehen und Ergebniswerte komparativ zu betrachten, werden die Entscheidungen der Anpassungstests der Vollständigkeit halber mit angegeben. Diese werden zum Signifikanzniveau $\alpha = 0,05$ durchgeführt und versucht, die Nullhypothese „Der untersuchte Datensatz ist Benford-verteilt“ abzulehnen. Es existieren für die ersten Ziffern neun und für die zweiten Ziffern zehn Merkmalsklassen. Die Ablehnung wäre genau dann durchzuführen, wenn die auf Basis der in Abschnitt 4.3 dargelegten Methodik ermittelten kritischen Werte durch die Testgrößen überstiegen würden. Um die jeweiligen Testgrößen bzw. Distanzmaße zu berechnen, ist es zunächst notwendig, die ersten beiden Stellen aus den Zahlen der erhobenen Datensätzen zu extrahieren und die Ziffernhäufigkeiten absolut und relativ zu bestimmen. Hierbei ist bei den zweiten Ziffern insbesondere darauf zu achten, dass diese auch tatsächlich vorhanden ist und nicht fälschlicherweise als „0“ ausgelesen wird. Dies wäre z.B. bei der Zahl 1,0 der Fall, da rechts von der „1“ an jeder weiteren Stelle eine „0“ steht. Somit besitzt die Zahl in Wahrheit nur eine Stelle und würde bei der Analyse der zweiten Ziffern nicht weiter berücksichtigt. Liegt hingegen die Zahl 1,07 vor, handelt es sich bei der zweiten Stelle um eine „echte“ null, die in der Häufigkeitszählung berücksichtigt werden müsste. Im Anschluss werden die hypothetischen Klassenwahrscheinlichkeiten auf Basis der Stichprobenumfänge ermittelt und mittels dieser sowie der empirischen Häufigkeiten die entsprechenden Werte der Testgrößen ermittelt.

Um den gleichgewichteten Einbezug aller Verfahren sicherzustellen erscheint es - wie im vierten Kapitel ermittelt - am effizientesten, die Ergebnisgrößen zwischen den zu vergleichenden Datensätzen relativ zu betrachten. Hierzu werden zwischen diesen Rangzahlen für die jeweiligen Testverfahren und Abstandsmaße in Abhängigkeit der Benford-Deviation ermittelt. Aus diesen wird dann der arithmetische Mittelwert bestimmt und so die Datensätze in eine effiziente Prüfreihefolge in Abhängigkeit der Wahrscheinlichkeit vorhandener Deviationen gebracht.

5.3 Ergebnisse der Untersuchung

5.3.1 Ergebnisse des aggregierten Datensatzes

Die Ziffernverteilung der ersten Ziffern des Gesamtdatensatzes im Vergleich zu der idealen Benford-Verteilung zeigt Abbildung 14. Durch diese wird bereits deutlich, dass der Gesamtdatensatz nahezu ideal mit der Benford-Verteilung übereinstimmt. Dies bezeugen auch die Ergebnisse der einzelnen Testverfahren, welche in Tabelle 40 zusammengefasst sind. Bei einem Gesamtstichprobenumfang von 216.078 Daten, beträgt der χ^2 -Wert gerade einmal 9,70, die Testgrößen des KS-Testes und des Kuiper-Testes liegen bei 0,0021 bzw. 0,9977, was in keinem der Fälle die Ablehnung der Nullhypothese zur Folge hat. Auch die weiteren Distanzmaße weisen allesamt sehr geringe Werte auf: So liegt die Größe χ^2/n beispielsweise rechnerisch lediglich bei einem Wert von 0,00004. Insgesamt spricht dies eindeutig dafür, dass im Falle der ersten Ziffern eine Benford-Verteilung grundsätzlich unterstellt werden kann.

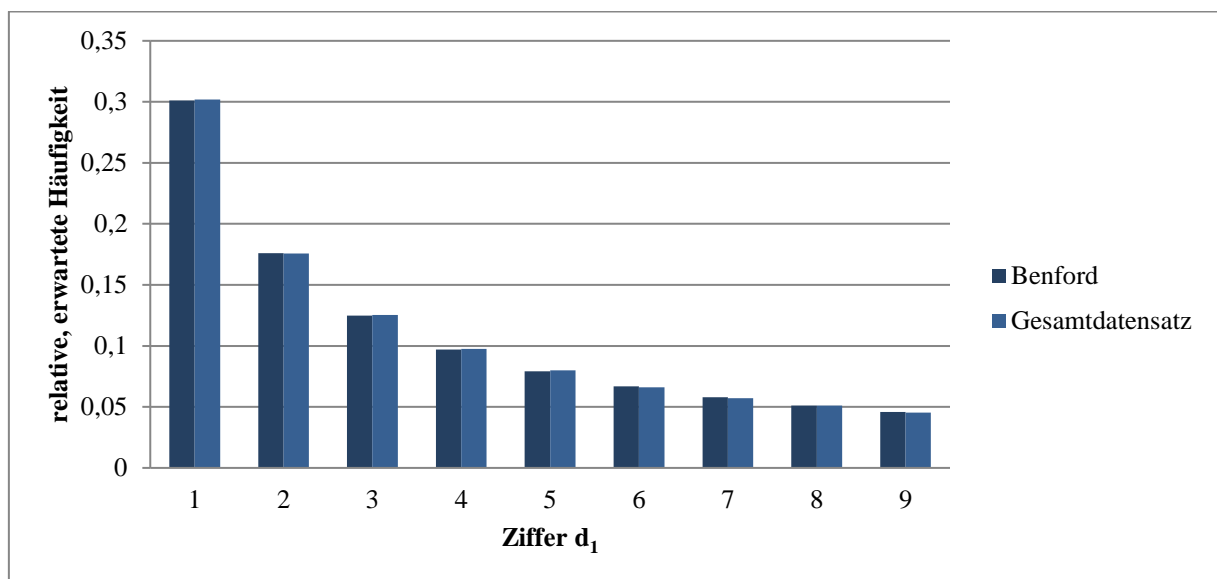


Abbildung 14: Verteilung der ersten Ziffern des Gesamtdatensatzes der historischen Daten

Die Ergebniswerte der zweiten Ziffern des Gesamtdatensatzes sind ebenfalls in Tabelle 40 dargestellt. Der unterschiedliche Stichprobenumfang im Vergleich zu den ersten Ziffern ist der Tatsache geschuldet, dass der Definition nach nicht jede erhobene Zahl auch eine „echte“ zweite Ziffer aufweist. So wurde beispielsweise bei einer glatten Zahl wie der 7,0 die zweite Ziffer als null nicht mit in die Untersuchung einbezogen, da dies andernfalls zu einer verzerrten Verteilung mit einer zu großen Anzahl an Nullen führen würde. Im Vergleich zu den Ergebnissen der ersten Ziffern fällt zunächst einmal auf, dass die Testgrößen in allen Fällen die Werte

der ersten Ziffern übersteigen. Alle drei Signifikanztests würden zudem die Nullhypothese statistisch signifikant ablehnen, da sowohl im Falle des χ^2 -Tests ($36,92 > 16,9$), des KS-Testes ($0,0031 > 0,0025$) sowie des Kuiper-Tests ($1,5031 > 1,321$) die Testgrößen jeweils im Ablehnungsbereich liegen. Demnach könnte der Verdacht entstehen, dass eine Benford-Verteilung für die zweiten Ziffern nicht unterstellt werden könnte.

	n	χ^2	KS	Kuiper	$\frac{\chi^2}{n}$	d*	a*	Manhattan	Tschebyscheff
Gesamtdatensatz 1. Ziffer	216.078	9,70	0,0021	0,9977	0,0000	0,0019	0,0014	0,0052	0,0010
Gesamtdatensatz 2. Ziffer	210.179	36,92*	0,0031*	1,5031*	0,0002	0,0045	0,0032	0,0108	0,0031

Tabelle 40: Ergebnisse des Gesamtdatensatzes der ersten und zweiten Ziffern

Abbildung 15 zeigt jedoch die Verteilung der zweiten Ziffern des Gesamtdatensatzes im Vergleich mit der idealen Benford-Verteilung. Zwar ist auch hier erkennbar, dass die Abweichung im Vergleich zu den ersten Ziffern geringfügig größer ist, jedoch kann grundsätzlich festgestellt werden, dass der Datensatz der Benford-Verteilung auch für die zweiten Ziffern sehr nahe kommt.

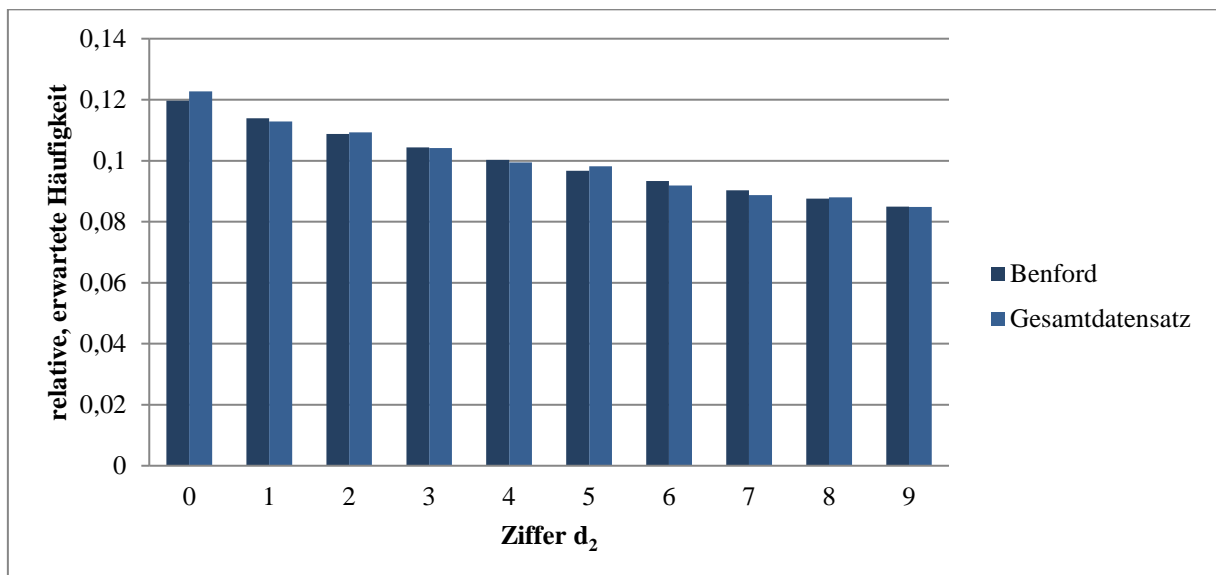


Abbildung 15: Verteilung der zweiten Ziffern des Gesamtdatensatzes der historischen Daten

Somit kann zusammengefasst festgehalten werden, dass der Datensatz sowohl für die ersten Ziffern als auch für die zweiten Ziffern insgesamt eine hohe Übereinstimmung mit der Benford-Verteilung aufweist und einer Anwendung des Benford-Testes auf historische Wirtschaftsdaten

somit argumentativ nichts im Wege steht, sodass im Folgenden sowohl die ersten als auch zweiten Ziffern untersucht werden. Da die Deviationen des Gesamtdatensatzes im Falle der zweiten Ziffern etwas größer ausfallen als bei den ersten Ziffern, sollte den Ergebnissen der ersten Ziffern im weiteren Untersuchungsverlauf eine höhere Aussagekraft zugesprochen werden.

5.3.2 Ergebnisse der einzelnen Staatsformen

Die Resultate des Benford-Testes der ersten Ziffern unterschieden nach den einzelnen Staatsformen in Form von Rangzahlen zeigt Tabelle 41. Die genauen Zahlenwerte können Tabelle 79 in Anhang III entnommen werden. Als Ergebnis ist festzustellen, dass die größte Benford-Deviation sehr eindeutig bei den Daten des NS-Regimes zu beobachten ist. Auf dem zweiten Platz liegen die Daten der Weimarer Republik, gefolgt von der Deutschen Demokratischen Republik. Die beste Datenqualität ist bei der Bundesrepublik Deutschland zu beobachten. Auffällig ist weiterhin, dass die Testergebnisse sehr robust durch nahezu alle Testgrößen bestätigt werden. Auch die errechneten Werte in Tabelle 79 in Anhang III zeigen deutlich, dass die Abweichungen der NS-Daten absolut am größten ausfallen: Während die Werte der BRD, DDR und Weimarer Republik in einer vergleichbaren Größenordnung liegen, fallen die Ergebniswerte des NS-Regimes zum überwiegenden Teil deutlich größer aus.

Rangzahlen 1. Ziffern	n	χ^2	KS	Kui- per	$\frac{\chi^2}{n}$	d*	a*	Man- hattan	Tscheby- scheff	Ge- samt
<i>Bundesrepublik Deutschland</i>	96.769	2	4	1	4	4	4	4	4	4
<i>DDR</i>	45.597	4	3	3	2	3	3	3	3	3
<i>Nationalsozialis- tisches Regime</i>	20.523	1	1	4	1	1	1	1	1	1
<i>Weimarer Republik</i>	53.189	3	2	2	2	2	2	2	2	2

Tabelle 41: Rangzahlen der Testgrößen der unterschiedlichen Staatsformen bei Analyse der ersten Ziffern

Die Ziffernverteilungen der ersten Ziffern der Staaten sind in Abbildung 16 dargestellt. Betrachtet man die Resultate des NS-Regimes so fällt auf, dass geringfügig zu wenig Sechsen und geringfügig zu viele Einsen ursächlich für die auffälligen Abweichungen sind. Konkrete Abweichungsmuster, die Rückschlüsse auf Ursachen der Abweichungen wie z.B. Rundungen geben, sind jedoch nicht zu erkennen.

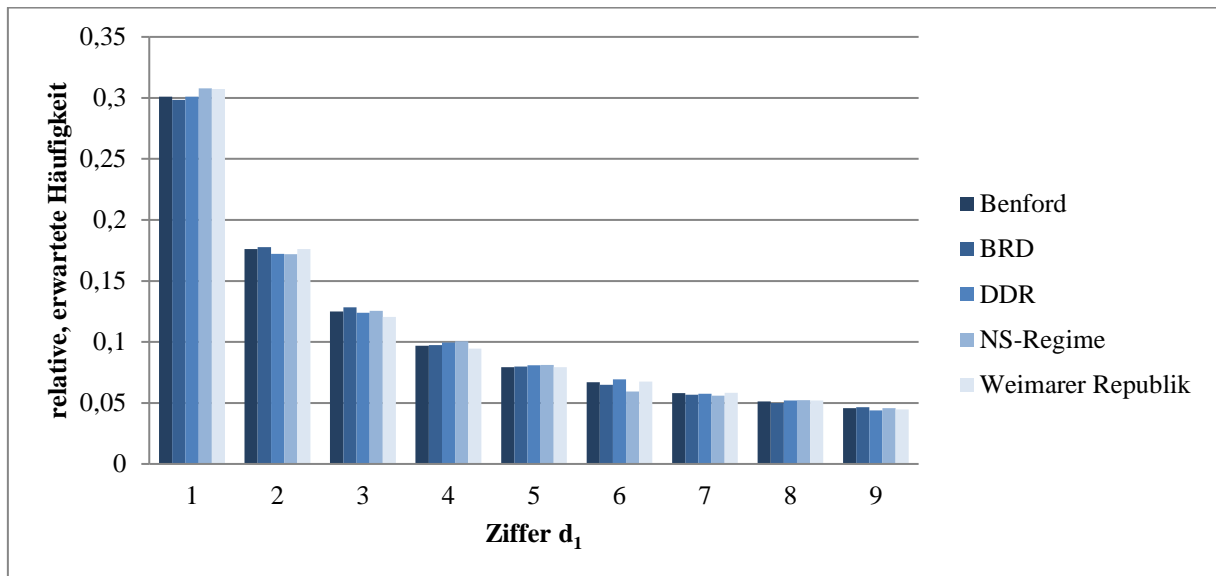


Abbildung 16: Ziffernverteilung der ersten Ziffern der einzelnen Staatsformen

Auch die Ergebnisse der Z-Statistik bestätigen die obigen Erkenntnisse mathematisch, wie Tabelle 42 belegt: Die größte Abweichung tritt im Falle des NS-Regimes mit einem Wert von 4,37 bei der Zahl sechs auf.

	Klasse mit größter Deviation	Wert der Z-Statistik
<i>Bundesrepublik Deutschland</i>	$d_1 = 6$	3,25
<i>DDR</i>	$d_1 = 2$	2,22
<i>NS-Regime</i>	$d_1 = 6$	4,37
<i>Weimarer Republik</i>	$d_1 = 3$	3,11

Tabelle 42: Ergebnisse der Z-Statistik der Staatsformen bei Analyse der ersten Ziffern

Die Rangzahlen der Testergebnisse und Abstandsmaße bei Analyse der zweiten Ziffern sind in Tabelle 43 zusammengefasst. Analog können die errechneten Werte Tabelle 80 in Anhang III entnommen werden. Hier weist die Weimarer Republik die größten Benford-Abweichungen auf. Auf dem zweiten Platz folgen das NS-Regime sowie die Deutsche Demokratische Republik. Die Daten der Bundesrepublik weisen erneut die beste Konformität mit der Benford-Verteilung auf. Auch hier sind die Ergebnisse bis auf kleinere Ausnahmen robust über alle Testverfahren anzusehen. Betrachtet man die Ergebniswerte in Tabelle 80 ergeben sich hierbei ähnliche Beobachtungen wie bei den ersten Ziffern: Auch hier liegen die Werte der erstplatzierten Staatsform – hier die Weimarer Republik – deutlich über denjenigen der drei übrigen.

Rangzahlen 2. Ziffern	n	χ^2	KS	Kui- per	$\frac{\chi^2}{n}$	d*	a*	Man- hattan	Tscheby- scheff	Ge- samt
<i>Bundesrepublik Deutschland</i>	94.576	2	4	2	3	4	4	4	4	4
<i>DDR</i>	44.600	3	2	3	3	3	2	3	3	2
<i>Nationalsozialis- tisches Regime</i>	19.453	4	3	4	2	2	3	2	2	2
<i>Weimarer Republik</i>	51.550	1	1	1	1	1	1	1	1	1

Tabelle 43: Rangzahlen der Testgrößen der unterschiedlichen Staatsformen bei Analyse der zweiten Ziffern

Die Häufigkeitsverteilung der zweiten Ziffern zeigt Abbildung 17. Deutlich zu erkennen ist, dass in dem Datensatz der Weimarer Republik ein deutlicher Überhang an Nullen im Vergleich zu der idealen Benford-Verteilung besteht. Des Weiteren ist die Ziffer Neun leicht unterrepräsentiert. Dieses Verteilungsmuster tritt im Allgemeinen bei systematischen Aufrundungen von Zahlen auf. Die restlichen Staatsformen zeigen insgesamt eine deutlich bessere Annäherung an die Benford-Verteilung. Auffälligkeiten sind hier nicht zu beobachten.

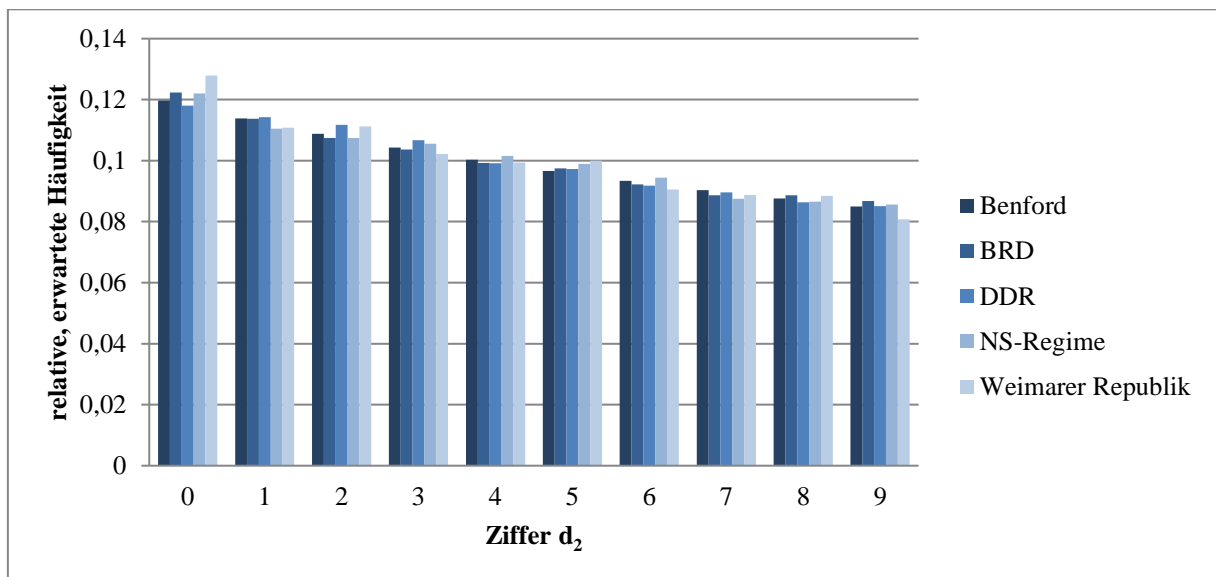


Abbildung 17: Ziffernverteilung der zweiten Ziffern der einzelnen Staatsformen

Die Ergebnisse der Z-Statistik für die zweiten Ziffern zeigt Tabelle 44. Auch hier werden die aus Abbildung 17 gewonnenen Erkenntnisse mathematisch untermauert. So tritt die eindeutig größte Abweichung bei der Weimarer Republik bei der Ziffer null auf. Der Ergebniswert liegt mit 5,47 deutlich über den maximalen Abweichungen der übrigen Staatsformen.

	Klasse mit größter Deviation	Wert der Z-Statistik
<i>Bundesrepublik Deutschland</i>	$d_2 = 0$	2,53
<i>DDR</i>	$d_2 = 2$	2,01
<i>NS-Regime</i>	$d_2 = 1$	1,49
<i>Weimarer Republik</i>	$d_2 = 0$	5,47

Tabelle 44: Ergebnisse der Z-Statistik der Staatsformen bei Analyse der zweiten Ziffern

Ergänzend zu den bisherigen Betrachtungen zeigt Tabelle 45 die Ergebnisse des Verzerrungsfaktormodells für die einzelnen Staatsformen, welche die bereits dargestellten Erkenntnisse weitestgehend bestätigen: Die Mittelwerte des NS-Regimes und der Weimarer Republik weisen die größten Benford-Deviations auf. Mit 0,57% und 0,51% Abweichung stimmen die Mittelwerte der BRD und der DDR nahezu ideal mit dem durch die Benford-Verteilung zu erwartenden Mittelwert überein. Das negative Vorzeichen bei allen Datensätzen signalisiert zudem, dass die tatsächlichen Mittelwerte in allen Fällen unterhalb des zu erwartenden Wertes liegen.

Ergebnisse Verzerrungsfaktor	n	VF	Rang
<i>Bundesrepublik Deutschland</i>	96.769	-0,0057	3
<i>DDR</i>	45.597	-0,0051	4
<i>NS-Regime</i>	20.523	-0,1268	1
<i>Weimarer Republik</i>	53.189	-0,0823	2

Tabelle 45: Ergebnisse des Verzerrungsfaktor-Modells der Staatsformen

Zusammengefasst bleibt festzuhalten, dass bei der Analyse der ersten Ziffern die Daten des NS-Regimes sehr deutlich die größte Benford-Deviation aufweisen. Während bei Betrachtung der Häufigkeitsverteilung keine speziellen Muster erkennbar waren, tritt bei den zweiten Ziffern im Falle der Weimarer Republik, welche hier die eindeutig größten Deviationen aufweist, ein Verteilungsmuster auf, welches typisch für systematische Aufrundungen ist: Dieses ist gekennzeichnet durch einen Überhang an Nullen und einem Defizit der Neunen. Die Bundesrepublik Deutschland zeigt sowohl im Falle der ersten als auch der zweiten Ziffern die beste Konformität mit der Benford-Verteilung.

5.3.3 Ergebnisse der Einzeljahre

Um der Fragestellung nachzugehen, ob im Zeitverlauf Unterschiede in der Datenqualität bei den verschiedenen Staatsformen zu beobachten sind, erfolgt im Folgenden die Anwendung des Benford-Testes auf die Einzeljahre. Hierbei werden die Ergebnisse der Bundesrepublik Deutschland und der Deutschen Demokratischen Republik in den Einzeljahren in einem Vergleich gegenübergestellt.

5.3.3.1 Ergebnisse der Einzeljahre der Weimarer Republik

Die Einzeljahrauswertung des Benford-Testes der ersten Ziffern der Wirtschaftsdaten der Weimarer Republik zeigt Tabelle 81 in Anhang IV. Diejenigen Jahre mit einer im Vergleich höheren Benford-Deviation wurden hierbei mittels breiter Schrift hervorgehoben. Namentlich sind dies die Jahre 1918, 1919, 1924, 1927 und 1931, wobei die Abweichungen nahezu konstant durch alle Testgrößen angezeigt werden. Auffällig ist hierbei insbesondere, dass auftretende Benford-Deviationsen ausschließlich in diesen genannten Jahren auftreten. Für die restlichen Jahre ergeben sich bei allen Testgrößen deutlich kleinere Werte. Dies lässt darauf schließen, dass lediglich die genannten auffälligen Jahre ursächlich für die zweithöchste Gesamtdeviation innerhalb der vier betrachteten Staatsformen sind.

Tatsächlich aussagekräftig sind die o.g. Ergebnisse jedoch nur, wenn der Benford-Test diese bei Analyse der zweiten Ziffern bestätigt. Eingesehen werden kann die Einzeljahrauswertung der zweiten Ziffern in Tabelle 82 in Anhang IV. Dort zeigt sich, dass die größten Abweichungen in den Jahren 1919, 1924, 1925/26 und 1931 auftreten. Auch hierbei fällt auf, dass die Testgrößen in diesen Jahren deutlich oberhalb derer der restlichen Jahre liegen und damit sehr eindeutig die insgesamt höchste Deviation implizieren.

Vergleicht man die auffälligen Jahre der Analysen der ersten und zweiten Ziffern miteinander, lässt sich feststellen, dass übereinstimmend in beiden Untersuchungen die Jahre 1919, 1924 und 1931 Auffälligkeiten zeigen. Demzufolge können die Abweichungen dieser Jahre als robust angesehen werden.

Abbildung 18 zeigt die Häufigkeitsverteilung der ersten Ziffern der als robust auffällig identifizierten Jahre 1919, 1924 und 1931. Ins Auge fallend sind hierbei lediglich die deutlichen Überhänge der Ziffern Eins und Zwei. Ansonsten sind hier, wie auch bei dem Gesamtdatensatz der Weimarer Republik, keine besonderen Verteilungsmuster zu erkennen.

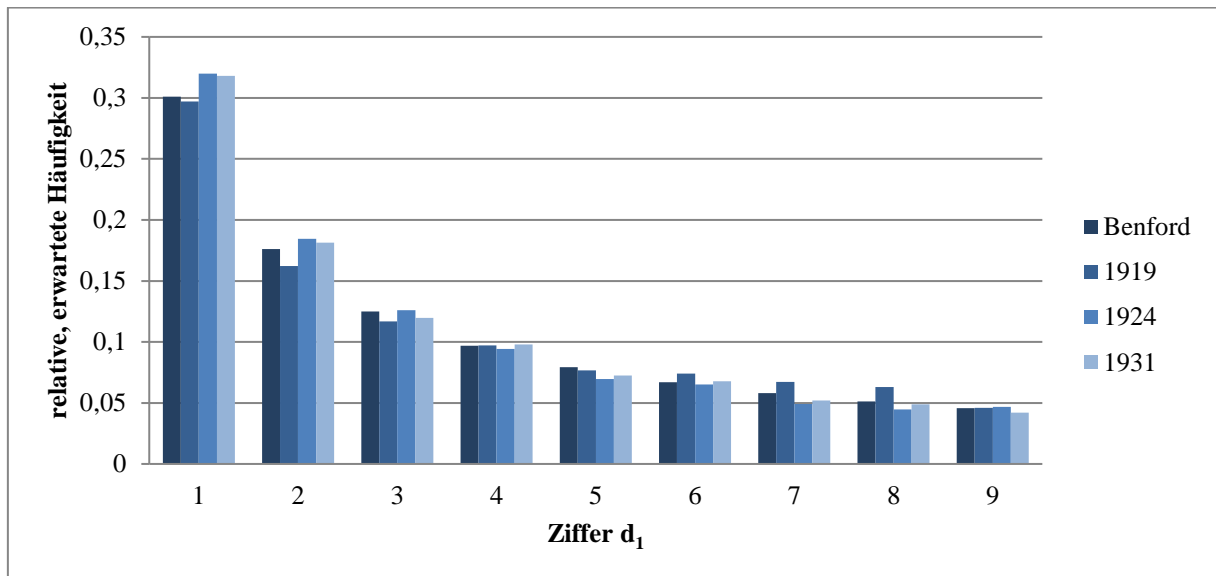


Abbildung 18: Verteilung der ersten Ziffern der auffälligen Jahre der Weimarer Republik

Abbildung 19 zeigt die Häufigkeitsverteilungen für die zweite Stelle der als auffällig indizierten Jahre. Wie schon bei dem Gesamtdatensatz der Weimarer Republik fällt auch hier der starke Überhang an Nullen in allen drei Jahren ins Auge. Die Neunen sind leicht unterrepräsentiert. Demzufolge deuten auch hier die Beobachtungen darauf hin, dass Daten dieser Jahre aufgerundet worden sein könnten. Weiterhin auffällig, jedoch ohne plausible Erklärung, ist ein deutlicher Überhang an Zweien in dem Jahr 1931.

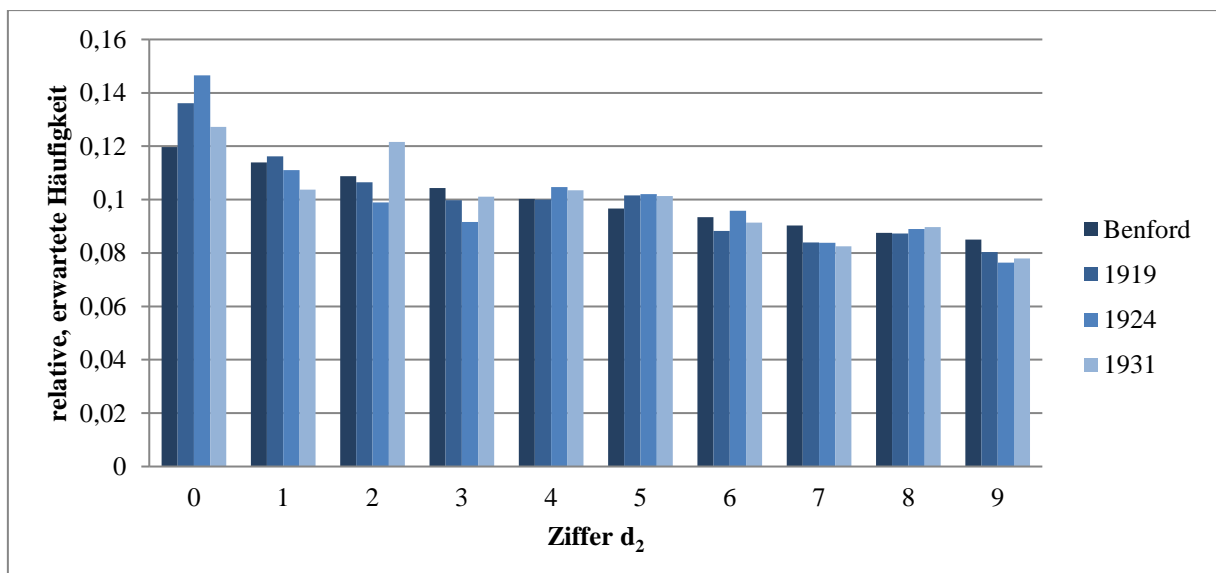


Abbildung 19: Verteilung der zweiten Ziffern der auffälligen Jahre der Weimarer Republik

Zusammenfassend bleibt festzuhalten, dass die Jahre 1919, 1924 und 1931 sowohl bei den ersten als auch bei den zweiten Ziffern die größten Abweichungen von der Benford-Verteilung

aufweisen. Im Falle der zweiten Ziffern deutet die Häufigkeitsverteilung zudem darauf hin, dass Daten Zahlen systematisch aufgerundet worden sein könnten. Ob diese Ergebnisse in dieser Form auch plausibel sind, wird in Abschnitt 5.4 näher betrachtet.

5.3.3.2 Ergebnisse der Einzeljahre des NS-Regimes

Für die Wirtschaftsdaten des NS-Regimes wurde ebenfalls eine Einzeljahrauswertung durchgeführt. Die Ergebnisse des Benford-Testes für die ersten Ziffern sind Tabelle 83 in Anhang V zu entnehmen. Von den acht betrachteten Zeitabschnitten wiesen die Jahre 1935, 1936, 1939 sowie 1941/42 die auffälligsten Ergebnisse auf, wobei sich die Resultate robust für sämtliche Testgrößen ergeben. Ähnlich wie bei den Zahlen der Weimarer Republik weisen die übrigen Jahre deutlich geringere Ergebnisgrößen auf, sodass davon ausgegangen werden kann, dass die Ziffernverteilungen hier der Benford-Verteilung entsprechen bzw. sehr nahe kommen. Im Umkehrschluss müssen demnach die als auffällig identifizierten Jahre ursächlich für die hohe Abweichung des Gesamtdatensatzes sein.

Tabelle 84 in Anhang V zeigt die Einzeljahreergebnisse des Benford-Testes bei Analyse der zweiten Ziffern. Bei diesen weisen die Jahre 1936, 1937, 1939 und 1941/42 die größten Deviationen auf. Hierbei liegen die Resultate der auffälligen Jahre jedoch nicht so deutlich über den übrigen Werten, wie es bei den Ergebnissen der ersten Ziffern der Fall ist. Dennoch liegen die Testergebnisse bis auf einzelne Ausnahmen bei dem Chi²- und Kuiper-Test sowie der Größe a^* für die angegebenen Jahre allesamt in der Spitzengruppe, sodass auch hier von stabilen Ergebnissen ausgegangen werden kann.

Vergleicht man die Ergebnisse der ersten und zweiten Ziffern miteinander, so stimmen drei der jeweils vier auffälligen Jahre überein: Die Jahre 1936, 1939 und 1941/42 weisen bei beiden Untersuchungen auffällige Ergebnisse auf, sodass diese als robuste Abweichungen angesehen werden können.

Die Häufigkeitsverteilungen der ersten und zweiten Ziffern der als auffällig identifizierten Jahre des NS-Regimes zeigen Abbildung 20 und 21. Für die ersten Ziffern ergibt sich hierbei bei allen Jahren ein Überhang an Einsen, wohingegen die Zahlen sechs, sieben und neun in den Jahren 1939 sowie 1941/42 leicht unterrepräsentiert sind. Diese Unregelmäßigkeit kann ein Indiz für systematische Aufrundungen im Datensatz sein.

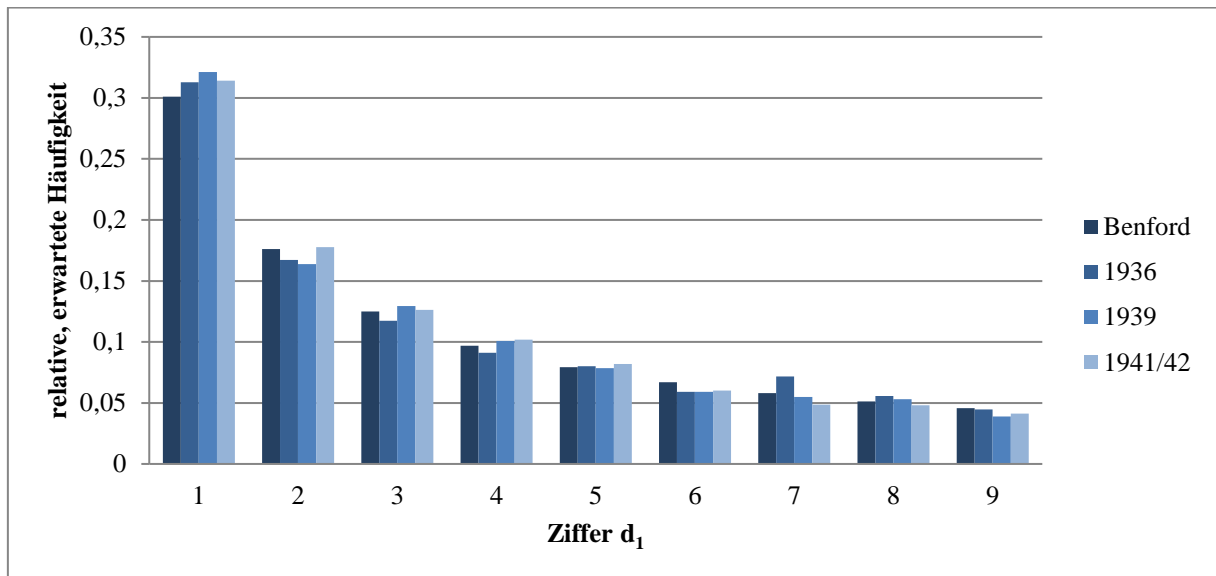


Abbildung 20: Verteilung der ersten Ziffern der auffälligen Jahre des NS-Regimes

Die Häufigkeitsverteilung der zweiten Ziffern zeigt Abbildung 21. Für die Jahre 1936 und 1939 fällt der deutliche Überhang an Nullen unmittelbar ins Auge, welcher jedoch im Jahr 1941/42 nicht auftritt. In den Jahren 1936 und 1939 sind ebenfalls geringfügig zu wenige Neunen zu beobachten, sodass auch hier im Ansatz ein Rundungsmuster zu erkennen ist. Für das Jahr 1941/42 kann diese Vermutung definitiv ausgeschlossen werden.

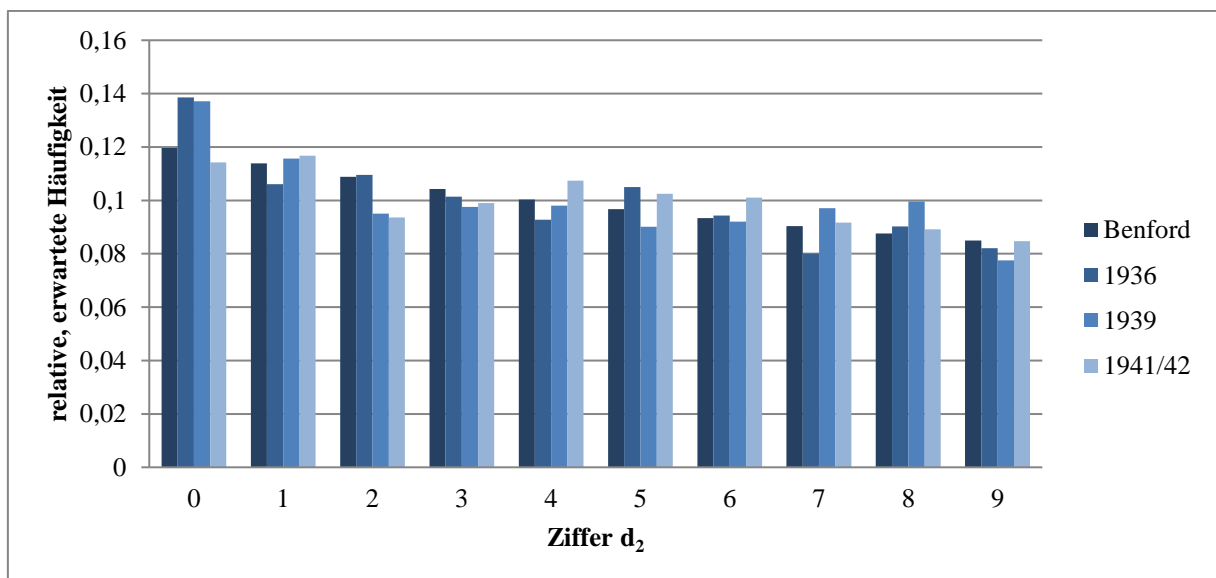


Abbildung 21: Verteilung der zweiten Ziffern der auffälligen Jahre des NS-Regimes

Im Falle der Wirtschaftsdaten des NS-Regimes zeigen v.a. die Jahre 1936, 1939 und 1941/42 auffällig hohe Abweichungen. Typische Verteilungsmuster, die auf systematisches Aufrunden der Zahlen schließen lassen, treten hierbei im Ansatz für die ersten Ziffern in den Jahren 1939

und 1941/42 auf, für die zweiten Ziffern in den Jahren 1936 und 1939. Auch diese Erkenntnisse werden in Abschnitt 5.4 im Rahmen der Bewertung einem grundlegenden Plausibilitätscheck unterzogen.

5.3.3.3 Ergebnisse der Einzeljahre der Deutschen Demokratischen Republik und der Bundesrepublik Deutschland

Die Ergebnisse des Benford-Testes bei Analyse der ersten Ziffern der Einzeljahre im direkten Vergleich zwischen der Bundesrepublik Deutschland und der Deutschen Demokratischen Republik zeigt Tabelle 85 in Anhang VI. Grundsätzlich bestätigen sich auch im Einzeljahresvergleich die in Abschnitt 5.3.2 gewonnenen Erkenntnisse hinsichtlich der höheren Benford-Deviation des Zahlenmaterials der Deutschen Demokratischen Republik. So weisen diese in 24 der insgesamt 35 vergleichbaren Jahre eine höhere Wahrscheinlichkeit vorhandener Manipulationen auf, was einem relativen Anteil von 69% entspricht.

Die größten Abweichungen im Falle der BRD treten hierbei in den Jahren 1958, 1959, 1964, 1981 und 1988 auf. In keinem der Jahre werden die Ergebnisse jedoch eindeutig durch alle Testverfahren bestätigt. So tritt die im Mittel größte Abweichung im Jahr 1964 auf, wobei jedoch keine der Testgrößen hierfür tatsächlich den größten Wert aufweist. Lediglich die Größe χ^2/n sowie die Tschebyscheff-Distanz indizieren jeweils einen zweiten Platz. Betrachtet man den Wert des Abstandsmaßes d^* , so liegt das Jahr 1964 bei diesem sogar lediglich auf dem 14. Platz. Ähnliche Beobachtungen können für die übrigen Jahre gemacht werden. Diese Auffälligkeit lässt darauf schließen, dass die Abweichungen der einzelnen Jahre in einem sehr ähnlichen Bereich liegen und sich nicht stark voneinander unterscheiden. Demzufolge kann nicht davon ausgegangen werden, dass die Deviationen in einem der Jahre als auffällig angesehen werden können.

Bei der Deutschen Demokratischen Republik sind die fünf Jahre mit den im Mittel größten Abweichungen die Jahre 1959, 1965, 1966, 1979 und 1990. Wenngleich auch hier in keinem der Jahre auffällige Deviationen durch alle Testverfahren angezeigt werden, fallen die Resultate derjenigen Jahre mit hohen Abweichungen hier deutlicher aus als bei der BRD. Betrachtet man hier die Daten aus 1979, das Jahr mit den insgesamt größten Deviationen, so zeigen hier sowohl die Größe χ^2/n als auch die Tschebyscheff-Distanz die insgesamt größten Abweichungen aller Jahre an. Immerhin fünf von acht Testgrößen ergeben Werte auf den ersten drei Plätzen. Nur

geringfügig bessere Werte liefert das Jahr 1990: Hier liefern vier von fünf Testgrößen Ergebnisse unterhalb der ersten drei Plätze.

Die Ergebnisse des Benford-Tests der Einzeljahre bei Anwendung auf die zweiten Ziffern zeigt Tabelle 86 in Anhang VI. Auch hier kann beobachtet werden, dass die Abweichungen in 25 von insgesamt 35 Jahren im Falle der DDR größer ausfallen als bei der Bundesrepublik Deutschland, was einem Anteil von rund 71% entspricht. In acht Jahren lagen die Werte der BRD über denen der DDR, in zwei Jahren konnte keine eindeutige Aussage getroffen werden. Wie bereits bei der Analyse der ersten Ziffern decken sich die Ergebnisse demnach auch hier mit den Resultaten aus Abschnitt 5.3.2.

Betrachtet man die Einzelergebnisse der BRD, so weisen hier die Jahre 1961, 1965, 1967, 1983 und 1989 die größten Abweichungen auf. Jedoch bestätigt sich hier ebenfalls das Bild der Analyse der ersten Ziffern: Insgesamt divergieren die Einzelplatzierungen der Testgrößen sehr stark und sprechen für insgesamt nicht eindeutige Ergebnisse. Auch hier muss davon ausgegangen werden, dass keines der Jahre auffällige Abweichungen aufweist. Dies bestätigt sich ebenfalls, wenn man diejenigen fünf Jahre mit der größten Abweichung zwischen den ersten und zweiten Ziffern vergleicht: Im Gegensatz zu den Ergebnissen der Weimarer Republik und des NS-Regimes stimmt hier keines der Jahre überein, was eindeutig gegen als robust interpretierbare Abweichungen spricht.

Der Benford-Test bei Anwendung auf die zweiten Ziffern des DDR-Datensatzes zeigt für die Jahre 1956, 1978, 1980, 1981 und 1983 die größten Abweichungen an, welche jedoch ebenfalls aufgrund der starken Heterogenität der Einzelergebnisse nicht als robust angesehen werden können. Wie bereits bei den Daten der BRD stimmt auch hier keines der fünf erstplatzierten Jahre der ersten und zweiten Ziffern überein. Aus diesem Grund kann auch hier kein Jahr identifiziert werden, was im Vergleich der übrigen Jahre deutlich erhöhte Abweichungen aufweist.

Zusammengefasst lässt sich somit festhalten, dass die Daten der Deutschen Demokratischen Republik im Vergleich zu denjenigen der BRD auch im Einzeljahresvergleich eine geringere Konformität und somit eine höhere Deviation mit bzw. von der Benford-Verteilung aufweisen. Jedoch lassen sich weder bei der DDR noch bei der BRD einzelne Jahre identifizieren, die stark auffällige Benford-Werte aufweisen. Im Fall der DDR lässt sich somit plausibel darauf schlie-

Ben, dass nicht einzelne Jahre mit einer sehr großen Deviation von der Benford-Verteilung ursächlich für die im Vergleich zur BRD höhere Abweichung sind, sondern vielmehr ein Großteil der analysierten Jahre eine geringe Abweichungen von dem Benfordschen Gesetz zeigen.

5.4 Bewertung der Ergebnisse

Zusammenfassend bleibt zunächst festzuhalten, dass die betrachteten makroökonomischen Daten der Benford-Verteilung sowohl für die ersten als auch für die zweiten Ziffern sehr gut folgen, sodass der Benford-Test hier uneingeschränkt zum Einsatz kommen kann. Damit die erhaltenen Ergebnisse als robust angesehen werden können, sollten bei Anwendung auf die ersten und zweiten Ziffern ein Großteil der Testgrößen und Maßzahlen die gleichen Resultate liefern.

Im Rahmen der Ergebnisinterpretation und -bewertung wird insgesamt davon ausgegangen, dass eine vorhandene Benford-Abweichung stark mit systematischen Manipulationen in dem analysierten Datensatz korreliert. Um zu prüfen, ob in diesem Kontext auch nichts gegen einen kausalen Zusammenhang spricht, werden die Ergebnisse des vorherigen Abschnitts im Folgenden hinsichtlich ihrer Plausibilität genauer untersucht.

Vergleicht man die Ergebnisse des Benford-Testes der betrachteten Staatsformen für die ersten und zweiten Ziffern kann beobachtet werden, dass die Bundesrepublik Deutschland sehr deutlich die höchste Benford-Konformität und damit die beste Datenqualität aufwies. Im Rahmen der Auswertung der Einzeljahre viel zudem kein Jahr durch erhöhte Benford-Deviations auf. Grundsätzlich repräsentiert dies exakt das Teilergebnis, was im Rahmen dieser Untersuchung auch zu erwarten war und kann als eine Art Negativ-Kontrollergebnis verstanden werden: Die Bundesrepublik Deutschland bildet als etablierte demokratische Staatsform kaum begründete Anhaltspunkte dafür, eigene Wirtschaftsdaten systematisch zu fälschen. Zum einen war die Bundesrepublik Deutschland in dem betrachteten Zeitfenster zwischen 1952 und 1990 nicht unmittelbar in kriegsrische Konflikte verwickelt, welche Motivation zu Datenfälschungen begründen würden. Des Weiteren war die wirtschaftliche Entwicklung der BRD in diesem Zeitraum sehr positiv: So betrug die jährliche wirtschaftliche Wachstumsrate bereits in dem ersten Jahrzehnt bis 1960 im Durchschnitt 8,2%. Bereits im Jahr 1970 betrug das reale BIP (in Preisen von 1991) 789 Mrd. Euro, was immerhin den rund 1,5-fachen Wert des BIPs von 1960 entspricht. Im letzten Jahrzehnt vor der Wiedervereinigung lag das durchschnittliche jährliche Wirtschaftswachstum bei 2,6%, was nur knapp unterhalb des Wertes von 3,6% liegt, welchen

die USA als weltweit stärkste Wirtschaftsnation in dem Zeitfenster zwischen 1947 und 2010 im Durchschnitt erzielte.¹⁹⁷ Demzufolge können die Ergebnisse des Datensatzes der Bundesrepublik Deutschland als plausibel angesehen werden und bestätigen, dass der Benford-Test ohne vorhandene Manipulationen auch keine Deviationen von der Benford-Verteilung attestiert.

Die zweitbeste Annäherung der untersuchten Staatsformen wies die Deutsche Demokratische Republik auf. Diese lag bei der geringfügig aussagekräftigeren Analyse der ersten Ziffern auf dem dritten Platz, bei den zweiten Ziffern teilte sich diese zusammen mit dem NS-Regime den zweiten Platz unter den Staatsformen. Zwar stach bei der Analyse der Einzeljahre keines der Jahre mit einer auffällig hohen Abweichung hervor, jedoch wurde im direkten Vergleich mit der Bundesrepublik Deutschland deutlich, dass die Abweichung der DDR-Zahlen bei einem Großteil der Einzeljahre höher ausfallen. Demzufolge kann darauf geschlossen werden, dass systematische Manipulationen nicht lediglich in Einzeljahren vorgenommen wurden, sondern - wenngleich in geringem Umfang - in dem politischen System stets allgegenwärtig waren. Auch dieses Ergebnis deckt sich grundsätzlich mit dem zu erwartenden Resultat: Eine langandauernde verheerende Arbeitsproduktivität, gekoppelt mit einem hohen Schuldenstand sowie geringen Investitionen in Infrastruktur und Wirtschaft sorgten für eine chronisch instabile Wirtschaftslage der Deutschen Demokratischen Republik.¹⁹⁸ Ingeheim standen die Regierungen der kommunistischen DDR jedoch stetig mit den Staaten der westlichen Welt, insbesondere der BRD, hinsichtlich der Fragestellung im Wettbewerb, welches politische System dem jeweils anderen überlegen ist. Aufgrund der objektiv schlechteren Lage des DDR-Regimes ist es demnach nur plausibel, dass Zahlen systematisch beschönigt worden sind. In der Folge ist auch dieses Ergebnis im Kontext als durchweg plausibel und nachvollziehbar zu bewerten.

Die Zahlen der Weimarer Republik lieferten insgesamt die zweithöchste Abweichung. Im Falle der ersten Ziffern belegte die erste demokratische Staatsform in der Geschichte Deutschlands den zweiten Platz, im Falle der zweiten Ziffern wies sie sogar die insgesamt größten Deviationen von der Benford-Verteilung auf. Da wie oben beschrieben den Ergebnissen der ersten Ziffern aufgrund der minimal höheren Aussagekraft im Zweifel jedoch die größere Bedeutung

¹⁹⁷ Vgl. Bundeszentrale für politische Bildung, Die wirtschaftliche Entwicklung der Bundesrepublik Deutschland, <http://www.bpb.de/politik/grundfragen/deutsche-verhaeltnisse-eine-sozialkunde/138634/wirtschaftliche-entwicklung?p=all> (Abgerufen am 22.12.2015), 2012.

¹⁹⁸ Vgl. Die Bundesregierung, Die wirtschaftliche Lage der DDR vor der Wende 1989, <http://www.bundesregierung.de/Content/DE/Magazine/MagazinWirtschaftFinanzen/072/s1-die-wirtschaftliche-lage-der-ddr-vor-der-wende.html> (Abgerufen am 28.03.2015), 2009.

zuzuschreiben ist, wird der Weimarer Republik in der Folge die insgesamt zweithöchste Abweichung attestiert. Möglicherweise hätte man nicht unbedingt vermutet, dass die Zahlen der Weimarer Republik eine höhere Abweichung von der Benford-Verteilung aufweisen, als die kommunistische Deutsche Demokratische Republik. Im Einzeljahresvergleich konnte jedoch festgestellt werden, dass die Jahre 1919, 1924 und 1931 auffällig hohe Deviationen von der Benford-Verteilung zeigen, welche sowohl durch die Analyse der ersten als auch der zweiten Ziffern bestätigt worden sind. Erkennbar war im Falle der Ziffernverteilung der zweiten Ziffern ein Muster für mögliche Aufrundungen in jedem der drei Jahre. Da die übrigen Jahre deutlich niedrigere Benford-Abweichungen zeigen, kann angenommen werden, dass die erhöhten Gesamtabweichungen vorwiegend auf diese drei auffälligen Jahre zurückzuführen sind. Die Abweichungen dieser auffälligen Jahre sind durch historische Geschehnisse jedoch sehr plausibel zu begründen: Im Jahr 1919 wurde mit dem Friedensvertrag von Versailles am 28. Juni der erste Weltkrieg offiziell beendet, nachdem im November 1918 die Kampfhandlungen im Zuge des Waffenstillstandsabkommens von Compiègne bereits eingestellt worden waren. Neben der eindeutigen Schuldzuweisung Deutschlands und seinen Verbündeten für den Kriegsausbruch waren territoriale Gebietsverluste des Deutschen Reiches ebenso Gegenstand des Vertrages wie Reparationsleistungen an die Siegermächte. Insgesamt führte dieses Maßnahmenpaket zu einer dramatischen Verschlechterung der Wirtschaftslage der noch jungen Weimarer Republik, weshalb der Versailler Vertrag durch die deutsche Regierung nur unter größtem Druck der Siegermächte unterzeichnet worden war.¹⁹⁹ Um schlagkräftige Argumente gegen zu hohe Entschädigungsleistungen vorbringen zu können, wäre es demnach eine denkbare Option, dass die Regierung der Weimarer Republik die wirtschaftliche Leistungsfähigkeit in dem Jahr 1919 gezielt schlechter darstellte, als dies eigentlich der Fall war. Lediglich das identifizierte Rundungsmuster kann in diesem Kontext nicht schlüssig erklärt werden. Auch das ebenfalls auffällige Jahr 1924, steht in unmittelbarem Zusammenhang mit einer vorausgegangenen wirtschaftlichen Krise: Im November 1923 fand die Inflation der Mark, auch bekannt als „Deutsche Hyperinflation“, ihren historischen Höhepunkt. Auslöser waren die immensen Kriegskosten des Deutschen Reiches in der Zeit von 1914 bis 1918 gekoppelt mit anschließenden Reparationsaufwendungen nach Kriegsende, welche zum größten Teil durch eine Neuverschuldung des Staates gedeckt wurden. Um den hieraus folgenden Zahlungsverpflichtungen nachzukommen, brachte die Regierung immer mehr Geld in den Umlauf, was zu einer dramatischen Geldentwertung

¹⁹⁹ Vgl. Bundeszentrale für politische Bildung, Versailles und Weimar, <http://www.bpb.de/apuz/30789/versailles-und-weimar?p=all> (abgerufen am 23.12.2015), 2008.

führte. Kostete ein Hühnerei am 09. Juni 1923 beispielsweise noch 800 Mark, waren es am 02. Dezember 1923 bereits 320 Mrd. Mark. Da die Reparationszahlungen nicht mehr geleistet werden konnten, kam es sogar zwischenzeitlich zu einer Besetzung des Rhein-Ruhr-Gebietes durch französische Truppen. Beendet wurde die Hyperinflation am 15. November 1923 mit Beginn der Währungsreform, welche mit der Einführung der Übergangswährung *Rentenmark* begann und im Oktober 1924 mit der Ablösung durch die neue Reichsmark beendet wurde.²⁰⁰ Diese Hyperinflation und dessen Umgang in der Statistik allein könnten jedoch ursächlich für eine Abweichung von der Benford-Verteilung sein. Problematisch ist die Tatsache, dass eine genaue und zuverlässige Datenerhebung, welche eine richtige Bewertung der Wirtschaftsleistungen der Höhe nach einschließt, unter ständig stark wachsenden Preisen äußerst schwierig ist: So hatten Waren- oder Dienstleistungen am Morgen des selbigen Tages zum Teil deutlich niedrigere Preise als am Abend. Doch welcher Preis ist im Zuge dessen korrekterweise anzusetzen, sofern eine Bewertung zu erfolgen hat? Da diese Frage äußerst schwer zu beantworten ist, kann kaum angenommen werden, dass Preisbewertungen in dem Jahr 1923 stets aktuell und korrekt erfolgten. Vielmehr wäre es nachvollziehbar, wenn Zahlen im Rahmen der Datenerhebung systematisch aufgerundet worden sind, um die enorme Inflation auszugleichen. Diese ungenauen Bewertungen könnten jedoch einen nicht unerheblichen Einfluss auf die Verteilungen der Anfangsziffern haben und auch das beobachtete Rundungsmuster gut erklären. 1931 ist das dritte Jahr der Weimarer Republik mit einer auffällig hohen Benford-Deviation und liegt inmitten der damaligen Weltwirtschaftskrise. Diese begann im Oktober 1929 mit dem sog. „Schwarzen Freitag“, einer der bis dato dramatischten Kursrutsche an der New Yorker Börse und weitete sich innerhalb kürzester Zeit insbesondere auch auf die Weimarer Republik aus. Ein Nachfragerückgang nach deutschen Produkten führte zu Überkapazitäten in der Produktion und damit zu Massenentlassungen der Arbeitnehmer. Innerhalb kürzester Zeit verdoppelte sich die Zahl der Arbeitslosen von 3 Mio. im Winter 1929/1930 auf 6,1 Mio. im Jahr 1932, was die ökonomische Abwärtsspirale weiter nach unten trieb. Der damalige Reichskanzler Heinrich Brüning reagierte daraufhin mit politischen Reaktionen, die die Krise weiter dramatisch zuspitzen sollten: Um die krisenbedingten Rückgänge der staatlichen Einnahmen zu kompensieren, erhöhte er die Einkommens- und Verbrauchssteuern und kürzte die Sozialleistungen, was in einer Deflation der Reichsmark und einen damit zusammenhängenden Exporteinbruch deutscher Produkte zur Folge hatte. Erschwerend kam hinzu, dass im Mai 1931 eine deutsch-österreichische Zollunion

²⁰⁰ Vgl. Spiegel-Online, Hyperinflation 1923 - Als die Mark vernichtet wurde, <http://www.spiegel.de/einestages/hyperinflation-1923-a-948427.html> (Abgerufen am 26.12.2015), 2009.

scheiterte, woraufhin sämtliche ausländische Kapitalanleger ihre Kredite zurückriefen.²⁰¹ Aufgrund dieser desaströsen ökonomischen Situation liegt es hier nahe, dass die Regierung durchaus Motivation für eine gezielte Anpassung der Wirtschaftsdaten hatte, um deren Handlungsmaßnahmen in Form der beschriebenen Deflationspolitik zu rechtfertigen. Dass die Zahlen hierbei systematisch aufgerundet worden sein könnten, ist hierbei ebenfalls eine durchaus als wahrscheinlich anzunehmende Option. In der Summe lässt sich folglich auch das Ergebnis der Weimarer Republik als plausibel bewerten: So scheinen in erster Linie drei Krisenjahre ursächlich für die große Benford-Deviation zu sein, welche aufgrund ihrer dramatischen Konsequenzen nachvollziehbare Gründe für systematische Manipulationen an Wirtschaftsdaten bzw. Abweichungen von der Benford-Verteilung mit sich bringen.

Die insgesamt größten Abweichungen traten im Falle des Nationalsozialistischen Regimes (1933 bis 1945) unter dem Diktator Adolf Hitler auf. Dieses belegte unter den betrachteten Staatsformen bei Analyse der ersten Ziffern einen sehr deutlichen ersten Platz, bei der Analyse der zweiten Ziffern lag das NS-Regime zusammen mit dem kommunistischen Regime der DDR auf dem zweiten Platz. Sehr deutliche Abweichungen wiesen hierbei sowohl für die ersten als auch für die zweiten Ziffern die Jahre 1936, 1939 und 1941/42 auf. Anzeichen für systematische Aufrundungen konnten hierbei bei den ersten Ziffern für die Jahre 1939 und 1941/42 sowie bei den zweiten Ziffern für die Jahre 1936 und 1939 beobachtet werden. Das Jahr 1936 stand insgesamt weitestgehend im Fokus der Vorbereitung auf den zweiten Weltkrieg. So erfolgte bereits 1935 die Transformation der Reichswehr in die Wehrmacht, was mit einer Aufstockung der Truppenstärke auf 550.000 Mann und der Wiedereinführung der Wehrpflicht verbunden war. Nahezu gleichzeitig erfolgte entgegen der Bestimmungen aus dem Versailler Vertrag die Wiedergründung der Deutschen Luftwaffe. Der zweite große Bruch des Friedensvertrages erfolgte dann bereits 1936 mit dem Einmarsch der Wehrmacht in die entmilitarisierte Zone des Rheinlands. Ebenfalls rief Hitler den sog. „Vierjahresplan“ auf, in welchem er die Einsatzfähigkeit und Kriegsbereitschaft der Wehrmacht und der Wirtschaft bis 1940 forderte.²⁰² Laut *Eichholtz* verzehnfachten sich hierdurch die Rüstungsausgaben von 1,5 Mio. Reichsmark im Jahr 1933 auf 11 Mio. Reichsmark im Jahr 1936 innerhalb von nur drei Jahren. Allein von 1935

²⁰¹ Vgl. Bundeszentrale für politische Bildung, Zerstörung der Demokratie 1930 – 1932, <http://www.bpb.de/geschichte/nationalsozialismus/dossier-nationalsozialismus/39537/zerstoerung-der-demokratie> (Abgerufen am 04.01.2016), 2011.

²⁰² Vgl. Bundeszentrale für politische Bildung, Der Weg in den Krieg, <http://www.bpb.de/geschichte/deutsche-geschichte/der-zweite-weltkrieg/199397/der-weg-in-den-krieg> (Abgerufen am 04.01.2015), 2015.

auf 1936 stiegen die Ausgaben um das Doppelte.²⁰³ Demzufolge liegt es nahe, dass das NS-Regime eigene Wirtschaftsdaten speziell in diesem Jahr nach oben anpasste, um den großen Anstieg der Rüstungsausgaben auch ökonomisch rechtfertigen zu können. Hierfür spricht auch das Verteilungsmuster, was zumindest bei den zweiten Ziffern auf systematische Aufrundungen hindeutet. Die beiden weiteren auffälligen Jahre 1939 und 1941/42 liegen historisch in der Zeit des zweiten Weltkrieges (1939-1945) und bilden die einzigen verfügbaren statistischen Daten des Deutschen Reiches aus dieser Zeitspanne. Zahlen aus den Jahren 1940, 1943, 1944 und 1945 standen nicht zur Auswertung zur Verfügung. Jedoch sind auch die Abweichungen in den betrachteten Kriegsjahren nicht grundlegend verwunderlich. So ist die Regierung einer Nation, welche sich im Kriegszustand befindet, sicherlich nicht zwingend bereit, statistische Daten über die eigene Leistungsfähigkeit korrekt anzugeben, da hierdurch auch die Feinde Rückschlüsse auf die wirtschaftliche Lage und daraus abgeleitet auch über die militärische Situation ziehen könnten. Sie würden damit völlig unfreiwillig mit Informationen versorgt. Des Weiteren könnte eine schlechtere Datenqualität auch dem Umstand geschuldet sein, dass der statistischen Genauigkeit der Datenerhebung während des Krieges nicht die höchste Bedeutung zu kommt und dieser deshalb weniger Aufmerksamkeit geschenkt wird. Nicht erklärt werden durch diese Ansätze lediglich die Verteilungsmuster, welche auch in diesen Jahren Rückschlüsse auf systematische Aufrundungen der Daten geben könnten. Hier sei jedoch angemerkt, dass diese Verteilungsmuster selbstverständlich auch randomisiert entstanden sein können und diese daher nicht zwingend begründet werden müssen. Demzufolge kann festgehalten werden, dass auch die Ergebnisse des NS-Regimes in der Form plausibel und nachvollziehbar sind.

Generell konnte somit dargelegt werden, dass sämtliche Ergebnisse aus dem Abschnitt 5.3 als weitestgehend plausibel erscheinen und der Benford-Test demnach seine Eignung zur Erkennung von Manipulationen bei der Anwendung auf makroökonomische Daten erneut unter Beweis gestellt hat. Somit erscheint auch die These schlüssig, dass eine hohe Benford-Abweichung in dem Themenbereich der makroökonomischen Wirtschaftsdaten durchaus einen brauchbaren Indikator für vorhandene Manipulationen in den untersuchten Datensätzen darstellt. Jedoch sei generell angemerkt, dass der Benford-Test auch hier keine Beweisfunktion erfüllt, sondern lediglich Anhaltspunkte für vorhandene Manipulationen geben kann. Indizier-

²⁰³ Vgl. Eichholtz, Dietrich, Geschichte der Deutschen Kriegswirtschaft 1939 – 1945, München 2003, S. 31.

ten Abweichungen sollten demnach zwingend durch ergänzende Prüfungsmaßnahmen nachgegangen werden und die Resultate so in jedem Fall weiter verifiziert werden, bevor auf deren Basis Anschuldigungen erhoben werden.

6. Anwendung des Benford's Law zur Manipulationsaufdeckung bei Kfz-Versicherungsschäden

6.1 Grundlagen des Datenmaterials

Die im Rahmen dieser Analyse verwendeten Datensätze wurden von der Inverso Gesellschaft für innovative Versicherungssoftware mbH in Kooperation mit der Versicherungskammer Bayern (VKB) zur Verfügung gestellt. Beinhaltet sind hierbei grundsätzlich die bei der VKB von den Versicherungsnehmern gemeldeten Kfz-Schadensfälle, die in den Jahren 2013 und 2014 bei dem Versicherungsunternehmen bearbeitet wurden. Da hierbei sämtliche zur Verfügung stehende Fälle der Schadensregulierung einbezogen worden sind, konnte auf eine randomisierte Datenselektion verzichtet werden.

Die Datensätze setzen sich aus zwei unterschiedlichen Teileinheiten zusammen: Zum einen sind dies sämtliche Daten derjenigen Schadensfälle, die in Partnerwerkstätten des Versicherungsunternehmens bearbeitet wurden. Dieser Datensatz wird im Folgenden als „Daten der Partnerwerkstätten“ bezeichnet. Der zweite Teil setzt sich aus denjenigen Fällen zusammen, die einer besonderen Prüfung durch Sachverständige der Versicherung unterzogen worden sind und im Folgenden als „Daten des Expertisechecks“ benannt werden.

Hierbei enthalten die Daten der Partnerwerkstätten insbesondere Informationen über das Datum des Auftrags, die beauftragte Partnerwerkstatt sowie deren Adresse und die Angabe, ob es sich bei dieser um eine Vertragswerkstatt eines Kfz-Herstellers oder aber um eine freie Werkstatt handelt. Hinsichtlich des schadhaften Kfz sind des Weiteren Angaben zum Hersteller und Modell, das Datum der Erstzulassung sowie das Kfz-Kennzeichen, die gezahlte Rechnungssumme sowie Informationen darüber enthalten, ob es sich um einen Kasko- oder aber um einen Haftpflichtschaden handelt. Die Daten wurden keiner speziellen Anonymisierung unterzogen, sodass kein Informationsverlust vorliegt, der Einfluss auf die Ergebnisse hätte.

Der zweite Datensatz des Expertisechecks beinhaltet insbesondere Informationen zu Monat und Jahr der Schadensentstehung, der Kreisstadt der Zulassung, dem Datum der Erstzulassung, der Angabe, ob es sich um einen Kasko- oder Haftpflichtschaden handelt, die Höhe des Bruttobetrag des Kostenvoranschlags vor und nach der Prüfung durch einen Sachbearbeiter der Versicherung sowie den im Anschluss tatsächlich an den Versicherungsnehmer ausgezahlten Bruttobetrag. Auf weitere Anonymisierungen wurde auch in diesem Datensatz verzichtet.

Von besonderem Interesse im Bezug zur Anwendung des Benford-Testes sind die in beiden Datensätzen enthaltenen Beträge der jeweiligen Schadensfälle. Da diese auf Basis der von den Kfz-Werkstätten ausgestellten Rechnungen oder Kostenvoranschlägen ermittelt werden, die durch die Versicherungsnehmer bei der Versicherung eingereicht werden, haben die Werkstätten und bei fiktiven Schadensabwicklungen auch die Versicherten allem Anschein nach ein ausgeprägtes Interesse, dass diese Beträge so hoch wie möglich ausfallen. Demnach ist es realistisch, dass Versicherungsnehmer in Absprache mit den jeweiligen Werkstätten Manipulationen an den ausgestellten Rechnungen vornehmen, um beidseitig von den sich daraus ergebenden Vorteilen zu profitieren. Als solche Manipulationen könnte z.B. auch die Durchführung zusätzlicher Reparatur- und Serviceleistungen verstanden werden, die nicht mit dem eigentlichen Schaden in Verbindung stehen, den Rechnungsbetrag jedoch erhöhen. Treten solche auf, müssten diese also durch den Benford-Test erkannt werden. Fraglich ist in diesem Kontext, bei welchen Schadensgruppen auffällig hohe Benford-Abweichungen identifiziert werden und ob diese plausibel begründet werden können. Im Falle des Expertisechecks sind weiterhin die Bruttobeträge der eingereichten Kostenvoranschläge jeweils vor und nach der Prüfung durch einen Sachbearbeiter von Interesse: Hier müsste die Datenqualität nach einer Prüfung grundsätzlich besser sein als bei den Beträgen vor der Prüfung, da falsche Angaben in den Rechnungen zumindest anteilig durch die zuständigen Mitarbeiter erkannt werden müssten. Funktioniert der Benford-Test im Falle von Kfz-Versicherungsschäden, müsste dieser hinsichtlich der dargelegten Punkte die korrekten Ergebnisse anzeigen.

6.2 Deskriptive Statistik und Untersuchungsdesign

Wie schon in der vorherigen Untersuchung dargelegt, ist es für die erfolgreiche Durchführung des Benford-Testes notwendig zu garantieren, dass die in Abschnitt 3.2 vorgestellten Voraussetzungen an den Datensatz eingehalten werden. Eine der dargelegten Anforderungen stellt das Nichtvorhandensein von festen Ober- und Untergrenzen innerhalb des Datensatzes dar, unter welchen auch Schwellenwerte verstanden werden können, bei deren Überschreitung die Angabe bzw. Erfassung der Daten erst erfolgt. Im Rahmen der durchgeführten Untersuchung wurden v.a. die Bruttobeträge des Reparaturaufwandes von Kfz-Schadensfällen betrachtet. Dies ist jedoch in Zusammenhang mit der geschilderten Voraussetzung aus zweierlei Gründen nicht gänzlich unproblematisch: Zum einen beinhalten viele Versicherungsverträge individuelle Ei-

genbeteiligungen der Versicherungsnehmer. Werden diese Beträge bei einem Unfall unterschritten, gibt es keinerlei Anlass für den Versicherungsnehmer, den entstandenen Schaden bei der Versicherung zu melden. Liegen also im Durchschnitt die Werte der Eigenbeteiligungen bei 300€ pro Schaden, würde die Verteilung der ersten Ziffern mit großer Wahrscheinlichkeit zu wenige Einsen und Zweien aufweisen. Es käme zu einer Abweichung von der Benford-Verteilung. Selbiges Phänomen könnte auch dadurch entstehen, dass bei der Einreichung eines Schadensfalls - in Abhängigkeit der Höhe des Schadens - die an die Versicherung zu entrichtenden Prämien ansteigen. Rein wirtschaftlich lohnt sich hierbei eine Meldung durch den Versicherungsnehmer erst ab der Überschreitung einer bestimmten Schadenshöhe. Beide Gegebenheiten könnten demnach wie eine virtuelle Untergrenze in dem Datensatz wirken, die zu Abweichungen bei der ersten Ziffern führen könnte. Auswirkungen auf die zweiten Ziffern sind hingegen nicht zu erwarten. Aus diesem Grund wird in Abschnitt 6.3.1 die Verteilung des Gesamtdatensatzes geprüft und somit getestet, ob die Datensätze insgesamt einer Benford-Verteilung folgen und der Eignung des Benford-Testes somit entsprochen wird.

Die weiteren genannten Voraussetzungen sind jedoch weitestgehend unproblematisch: Die Angabe der Daten erfolgt in allen Fällen in der Einheit Euro, sodass grundsätzlich die gleiche Dimension der Einheiten vorliegt. Somit sind auch keine Umrechnungen des Skalenniveaus notwendig. Da die Daten real erhobene Schadensfälle repräsentieren, ist zudem nicht von einer Zuweisung der Zahlen auszugehen, sodass der zufällige Ursprung der Daten gewährleistet ist. Bei den Zahlen handelt es sich zudem stets um absolute Euro-Beträge, sodass hierbei auch keine statistischen Kennzahlen vorliegen. Die betrachteten Zahlen bestehen insgesamt aus dem Gesamtbetrag der eingereichten Rechnungen bzw. Kostenvoranschlägen durch die Versicherungsnehmer, welche auf Basis der Einzelpositionen einer Reparatur errechnet wurden. Die hierbei durchgeführten mathematischen Operationen in Form von Additionen und Subtraktionen wirken sich jedoch nicht auf die Benford-Verteilung aus, sodass dies an und für sich nicht problematisch ist.

Einen Überblick über den Gesamtstichprobenumfang und einige statistische Kennzahlen der Datensätze gibt Tabelle 46. Damit von einer Benford-Verteilung ausgegangen werden kann ist es empfehlenswert, dass der Mittelwert größer als der Median ist und die Schiefe größer als Null, die Verteilung also rechtsschief ist. Dies impliziert, dass mehr kleine als große Zahlen in

dem Datensatz vorhanden sind, was eine Benford-Verteilung prinzipiell begünstigt. Im vorliegenden Fall ist dies stets der Fall, sodass der Unterstellung einer Benford-Verteilung nichts im Wege steht.

Wie zuvor beschrieben wurde, ist es zudem notwendig, dass eine vorgegebene Mindestgröße des Stichprobenumfanges vorliegt, damit die vorgestellten Testverfahren angewendet werden können. Notwendig sind z.B. für den Chi²-Anpassungstest bei der Analyse der ersten Ziffern 110 Daten. Sollen die zweiten Ziffern untersucht werden, stellen 59 Zahlen das Minimum dar. Empfehlenswert für aussagekräftige Ergebnisse sind - wie in Abschnitt 4.3 gezeigt - ca. 300 Werte. Betrachtet man die Stichprobengrößen der Datensätze, so sind diese Werte deutlich überschritten, sodass der Benford-Test grundsätzlich Anwendung finden kann. Nicht empfehlenswert erscheint lediglich ein Vergleich zwischen den beiden Datensätzen: So liegen für den Datensatz der Partnerwerkstätten 3.828 bzw. 5.108 Zahlen vor, für den Datensatz des Expertisechecks aber 51.380 bzw. sogar 63.245 Zahlen. Möchte man sämtliche in Abschnitt 4.3 analysierten Verfahren verwenden, wäre hier die Voraussetzung an ungefähr ähnliche Stichprobenumfänge nicht mehr erfüllt. Innerhalb der Datensätze sind komparative Auswertungen jedoch ohne Einschränkungen möglich. Wird der Datensatz weiter untergliedert, müssten beide Voraussetzungen jedoch bei jedem sich ergebenden Teildatensatz stets neu überprüft werden.

Datensatz	Stichproben- umfang <i>n</i>	Mittelwert	Median	Schiefe
<i>Partnerwerkstätten</i>				
2013	3.828	2.859,42	2.302,81	3,09
2014	5.108	2.743,92	2.192,26	2,72
<i>Expertisecheck</i>				
2013	51.380	2.982,75	2.291,70	5,83
2014	63.245	2.908,59	2.290,85	3,78

Tabelle 46: Stichprobenumfang und statistische Kennzahlen der untersuchten Datensätze der Kfz-Unfallschäden

Ideal wäre es dem Anschein nach wenn es möglich wäre, den Benford-Test auf die eingereichten Einzelschäden anzuwenden. Dies würde es zeitgleich ermöglichen, im Falle entdeckter Abweichungen Rückschlüsse auf enthaltene Manipulationen zu ziehen und damit in jedem Einzelfall Betrugereien zu entlarven. Für die Versicherungen würde dies mit einer enormen Aufwandsersparnis einhergehen, da die Täuschungsversuche gezielt aufgedeckt werden könnten. Für

diese Vorgehensweise wäre es jedoch notwendig, dass für jeden Einzelschaden mehr als einhundert Einzelpositionen vorliegen müssten. Da jedoch jeder Schadensfall lediglich 1-3 Zahlen liefert, muss von dieser Vorgehensweise abgesehen werden.

Als eine sinnvolle alternative Prozedur erscheint es jedoch, Schadensfälle hinsichtlich bestimmter Untersuchungskriterien zusammenzufassen und auf Basis dessen Erkenntnisse darüber zu gewinnen, ob im Falle bestimmter Schadensgruppen die Wahrscheinlichkeiten für vorhandene Manipulationen höher ausfallen als bei anderen. Zu diesem Zweck werden die Datensätze zunächst hinsichtlich der folgenden Kriterien untergliedert: Der Datensatz der Partnerwerkstätten wird hinsichtlich der Punkte

1. Vertragswerkstatt oder freie Werkstatt
2. Kfz-Hersteller (Automarke)
3. Datum der Erstzulassung
4. Haftpflicht- oder Kaskoschaden

genauer untersucht. Bei den Daten des Expertisechecks werden die Kriterien

1. Schadenshöhe vor bzw. nach Prüfung durch Sachverständige sowie tatsächlich gezahlter Rechnungsbetrag
2. Datum der Erstzulassung

näher betrachtet. Unterscheidungen weiterer Untersuchungskriterien wie z.B. nach dem PLZ-Bereich der Werkstätten können nicht durchgeführt werden, da in diesem Fall die Untergliederungen zu klein würden und im Umkehrschluss die einzelnen Gruppen zu geringe Stichprobenumfänge aufweisen würden.

Die größten Auswirkungen im Falle gezielter Falschangaben und Manipulationen werden durch eine Fälschung der ersten Ziffern erreicht, weshalb der Benford-Test grundsätzlich auf diese angewendet werden sollte. Ist dies nicht möglich, z.B. weil der Gesamtdatensatz für die ersten Ziffern wider der Annahme keine Benford-Verteilung aufweist, sollte lediglich auf die zweiten Ziffern zurückgegriffen werden. Dies wird in Abschnitt 6.3.1 geprüft und die Untersuchung entsprechend fortgesetzt. Angewendet werden hierbei sämtliche in Abschnitt 3.4 vorgestellten und in Abschnitt 4.3 untersuchten Verfahren, wobei die Z-Statistik aufgrund ihrer Eignung lediglich dazu eingesetzt wird, diejenigen Klassen mit der größten Deviation zu identifizieren. Auch das Verzerrungsfaktormodell wird lediglich als ergänzendes Verfahren angewendet.

Zunächst werden die absoluten und relativen Häufigkeiten aller Merkmalsklassen ermittelt. Zu diesem Zweck werden die ersten beiden Ziffern einer jeden Zahl ausgelesen und gezählt. Im Anschluss erfolgt unter Verwendung der statistischen Testverfahren der Vergleich mit den Erwartungshäufigkeiten, die auf Basis der Stichprobenumfänge und der hypothetischen Klassenswahrscheinlichkeiten ermittelt werden. Sämtliche Ablehnungsbereiche werden auf Basis der in Abschnitt 4.3 ermittelten Vorgehensweisen errechnet und angegeben. Alle Anpassungstests werden hierbei zum Signifikanzniveau $\alpha = 0,05$ durchgeführt. Die Nullhypothese lautet für sämtliche Verfahren einheitlich $H_0 :=$ „Der untersuchte Datensatz ist Benford-verteilt.“ Es treten insgesamt neun Merkmalsklassen für die ersten Ziffern und zehn Merkmalsklassen für die zweiten Ziffern auf. Liegen die Testgrößen innerhalb der Ablehnungsbereiche, wäre die Nullhypothese abzulehnen. Gemäß der aus Kapitel 4 gewonnenen Erkenntnisse, ist diese Vorgehensweise im Rahmen der durchgeführten Untersuchung jedoch nicht empfehlenswert und lediglich von nebensächlicher Bedeutung: Die Signifikanztests werden ihrem Ursprung nach nur aus Gründen der Vollständigkeit durchgeführt und auf deren Interpretation weitestgehend verzichtet. Eine alternative Prozedur stellt wie bereits in der ersten empirischen Analyse der gleichwertige Einbezug sämtlicher Verfahren in die Schlussfolgerungen dar, bei welcher die Ergebnisse der Testgrößen einzelner Untersuchungsgruppen relativ zu den anderen Resultaten betrachtet werden. Auf Basis der erhaltenen Testgrößen werden jeweils Rangzahlen in Abhängigkeit von der Benford-Deviation errechnet. Aus diesen wird dann ein mittlerer Rang für jede der untersuchten Ausprägungen ermittelt, wobei alle Testverfahren und Maßzahlen gleich gewichtet werden. Durch die Bildung dieser effizienten Prüfreihenfolgen können dann Aussagen über die Wahrscheinlichkeit vorhandener Manipulationen getroffen werden, welche im Anschluss auf ihre Sinnhaftigkeit überprüft werden.

6.3 Ergebnisse der Untersuchung

6.3.1 Ergebnisse der Gesamtdatensätze

Abbildung 22 zeigt die Verteilung der ersten Ziffern der analysierten Gesamtdatensätze im Vergleich mit der idealen Benford-Verteilung. Zu erkennen ist, dass eine erkennbare Abweichung von dieser für alle Datensätze zu beobachten ist. Ursächlich hierfür erscheint auf dem ersten Blick eine deutlich zu hohe Anzahl vorhandener Einsen und Zweien. Im Umkehrschluss sind die Ziffern fünf bis neun unterrepräsentiert.

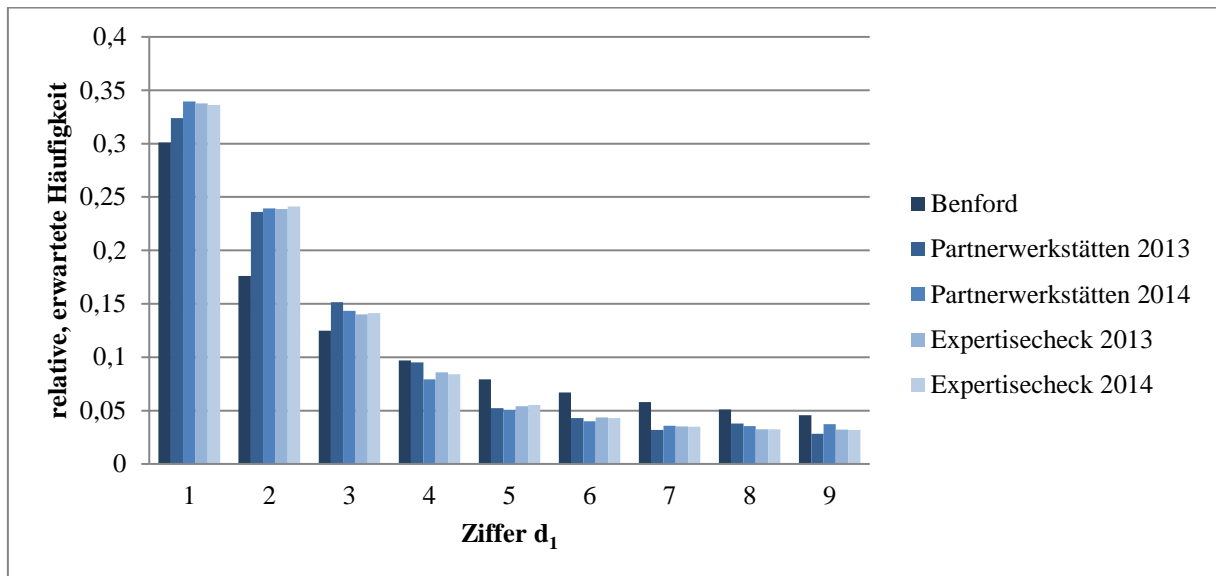


Abbildung 22: Verteilung der ersten Ziffern der Gesamtdatensätze

Diese Erkenntnis wird auch durch die Ergebnisse der Z-Statistik für die ersten Ziffern bestätigt, die in Tabelle 47 dargestellt sind. Im Falle aller Datensätze weist die Zahl zwei die größten Deviationen auf, wobei der Wert der Z-Statistik in Relation zu den restlichen Werten deutlich erhöht ist.

Z-Statistik der 1. Zif.	$d_1 = 1$	$d_1 = 2$	$d_1 = 3$	$d_1 = 4$	$d_1 = 5$	$d_1 = 6$	$d_1 = 7$	$d_1 = 8$	$d_1 = 9$
Partnerwerk- stätte									
2013	3,07	9,74	4,95	0,35	6,14	5,87	6,88	3,69	5,16
2014	5,97	11,83	3,95	4,23	7,61	7,70	6,81	5,07	2,83
Expertise- check									
2013	18,17	37,31	10,42	8,51	21,12	21,26	22,12	19,14	14,66
2014	19,22	43,04	12,47	11,06	22,18	24,18	25,00	21,33	16,66

Tabelle 47: Ergebnisse der Z-Statistik für die ersten Ziffern der Gesamtdatensätze

Eine plausible Ursache für die starken Abweichungen der Datensätze von der Benford-Verteilung könnten die zwischen dem Versicherungsnehmer und der Versicherung vereinbarten Eigenbeteiligungen oder aber die Konsequenz sein, dass eingereichte Schadensfälle die individuelle Risikostufe und damit die zu entrichtenden monatlichen Beiträge des Versicherungsnehmers erhöhen. Beides wirkt im Endeffekt auf die Verteilung wie eine Art Untergrenze, da sich die Meldung eines Schadens erst ab Überschreitung einer bestimmten Schadenshöhe für den

Versicherten rechnen. Demnach werden kleinere Schäden schlichtweg nicht gemeldet, was zu einer Abweichung von der Benford-Verteilung führt: Liegt der theoretische Schwellwert z.B. bei 1.200 €, so beinhaltet der Datensatz mit sehr großer Wahrscheinlichkeit zu viele Zweien für die ersten Ziffern, da Schäden erst ab Überschreitung dieses Schwellenwertes vermehrt gemeldet werden.

Vergleicht man hiermit die Verteilung der zweiten Ziffern der Gesamtdatensätze in Abbildung 23, lässt sich hierbei eine deutlich bessere Annäherung an die Benford-Verteilung beobachten. Ausreißer mit einer auffälligen Abweichung lassen sich hierbei bei keiner Ziffer identifizieren. Grundsätzlich sollten demnach auch die Testgrößen eine deutlich geringere Abweichung im Vergleich zu den ersten Ziffern attestieren.

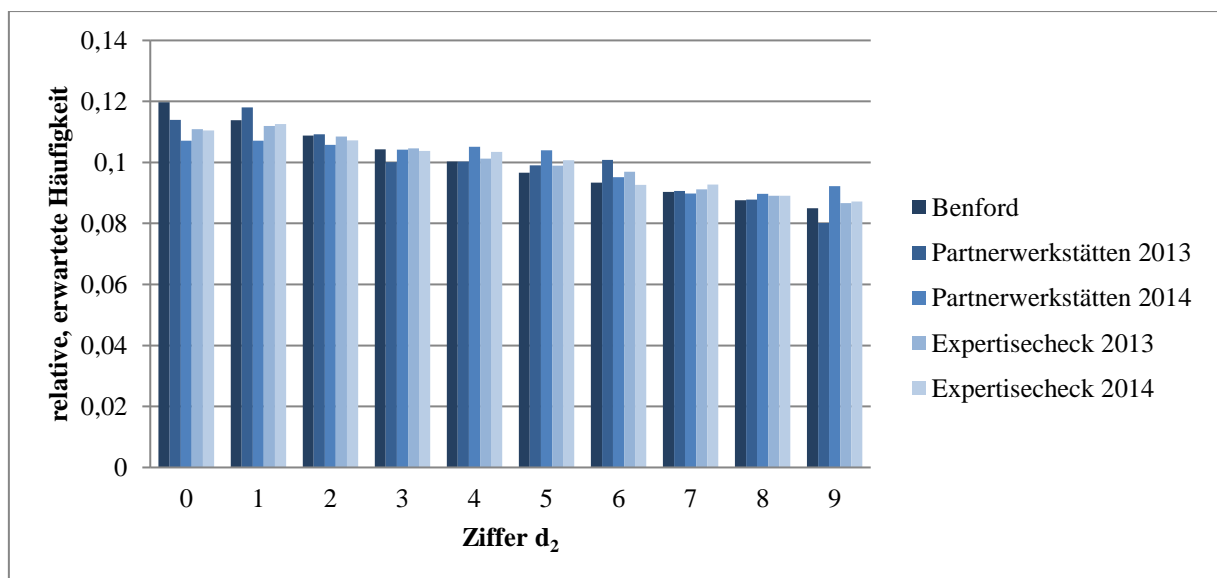


Abbildung 23: Verteilung der zweiten Ziffern der Gesamtdatensätze

Die Ergebnisse der statistischen Kennzahlen für die ersten Ziffern sind in Tabelle 48 dargestellt. Hierbei signalisieren die mit einem (*) markierten Werte jeweils eine signifikante Ablehnung der Nullhypothese, welche bei sämtlichen Datensätzen zu beobachten ist. Auch in der Höhe deuten die Ergebniswerte auf eine klare Rejektion der Benford-Verteilung hin, sodass sich die Vermutungen bei Betrachtung der Häufigkeitsverteilung sowie der Ergebnisse der Z-Statistik insgesamt bestätigen.

1. Ziffern	n	χ^2	KS	Kuiper	$\frac{\chi^2}{n}$	d*	a*	Manhattan	Tschebyscheff
Partnerwerkstätten									
2013	3.828	258,42*	0,1095*	6,795*	0,0675	0,0825	0,0916	0,2191	0,0601
2014	5.108	356,83*	0,1199*	8,592*	0,0699	0,0889	0,0926	0,2399	0,0631
Expertise-check									
2013	51.383	3.386,63*	0,1147*	26,017*	0,0659	0,0856	0,0960	0,2294	0,0627
2014	63.250	4.316,90*	0,1167*	29,357*	0,0683	0,0871	0,0967	0,2333	0,0652

Tabelle 48: Ergebnisse der statistischen Testgrößen der Gesamtdatensätze für die ersten Ziffern

Tabelle 49 zeigt die Ergebnisse der statistischen Kennzahlen für die zweiten Ziffern der Datensätze. Da sowohl für die ersten als auch für die zweiten Ziffern identische Stichprobenumfänge vorliegen, lassen sich deren Ergebnisse ideal vergleichen. Zwar führen die Testverfahren auch hier in acht Fällen zu einer statistisch signifikanten Ablehnung der Nullhypothese, jedoch sollte dieses Ergebnis nicht überinterpretiert werden: Vor allem im Falle des als stichprobeninvariant identifizierten KS-Tests werden die kritischen Werte für die Datensätze der zweiten Ziffern nur sehr knapp überschritten. Die χ^2 -Werte der Datensätze des Expertisechecks liegen ebenfalls mit 48,29 und 73,64 bei mehr als 51.000 bzw. 63.000 Zahlen in einem sehr guten Bereich. Demzufolge erscheint die Benford-Verteilung im Falle der zweiten Ziffern durchaus unterstellbar zu sein.

2. Ziffern	n	χ^2	KS	Kuiper	$\frac{\chi^2}{n}$	d*	a*	Manhattan	Tschebyscheff
Partnerwerkstätten									
2013	3.828	5,87	0,0058	0,6563	0,0015	0,0128	0,0019	0,0297	0,0075
2014	5.108	16,84	0,0227*	1,6245*	0,0033	0,0194	0,0272	0,0463	0,0126
Expertise-check									
2013	51.380	48,29*	0,0111*	2,5166*	0,0009	0,0106	0,0141	0,0222	0,0088
2014	63.245	73,71*	0,0126*	3,1625*	0,0012	0,0117	0,0148	0,0265	0,0092

Tabelle 49: Ergebnisse der statistischen Testgrößen der Gesamtdatensätze für die zweiten Ziffern

Aus diesen Erkenntnissen lässt sich schließen, dass die Anwendung des Benford-Testes auf die ersten Ziffern im Fall der Versicherungsdaten nicht möglich ist. Vielmehr muss der Betrachtungshorizont auf die zweiten Ziffern eingeschränkt werden. Da diese jedoch in den Gesamtdatensätzen eine gute Benford-Verteilung widerspiegeln, sollte diese Einschränkung keine allzu starken Konsequenzen haben und der Aussagekraft des Benford-Testes nichts im Wege stehen. Aus diesem Grund wird der Benford-Test im Folgenden lediglich auf die zweiten Ziffern der beiden Datensätze der Versicherungsschäden angewendet.

6.3.2 Ergebnisse des Datensatzes der Partnerwerkstätten

Eine erste Untergliederung des Datensatzes der Partnerwerkstätten erfolgt anhand der Art der Werkstatt, in welchem der eingereichte Schadensfall bearbeitet wurde. Hierbei sind die Partnerwerkstätten in die freien Werkstätten sowie die Vertragswerkstätten der Kfz-Hersteller, im Folgenden als „Markenwerkstätten“ bezeichnet, unterteilt. Die Ergebnisse des Benford-Testes der einzelnen Verfahren in Form von Rangzahlen bei Analyse der zweiten Ziffern zeigt Tabelle 50. Hierbei fallen die Ergebnisse sehr eindeutig aus: Sämtliche Verfahren in beiden Betrachtungsjahren 2013 und 2014 zeigen im Falle der freien Werkstätten eine höhere Benford-Abweichung an. Vergleicht man die Ergebnisse der Einzelgrößen untereinander in Tabelle 87 in Anhang VII so fällt auf, dass der Unterschied zwischen den Ergebniszahlen v.a. im Jahr 2013 sehr deutlich ausfällt.

Rangzahlen 2. Ziffern	n	χ^2	KS	Kuiper	$\frac{\chi^2}{n}$	d*	a*	Manhattan	Tschebyscheff	Gesamt
2013										
Markenwerkstätten	2.554	2	2	2	2	2	2	2	2	2
Freie Werkstätten	1.278	1	1	1	1	1	1	1	1	1
2014										
Markenwerkstätten	3.463	2	2	2	2	2	2	2	2	2
Freie Werkstätten	1.646	1	1	1	1	1	1	1	1	1

Tabelle 50: Rangzahlen der Testergebnisse der zweiten Ziffern des Datensatzes der Partnerwerkstätten bei Unterscheidung nach Art der Werkstatt

In Ergänzung hierzu bestätigen auch die Ergebnisse des Verzerrungsfaktormodells in Tabelle 51 diese Resultate: Sowohl in 2013 als auch in 2014 fallen die Abweichungen zu dem Mittelwert der Benford-Verteilung bei den freien Werkstätten größer aus.

Ergebnisse Verzerrungsfaktor	VF	Rang
2013		
<i>Markenwerkstätten</i>	-0,1161	2
<i>Freie Werkstätten</i>	-0,1562	1
2014		
<i>Markenwerkstätten</i>	-0,1240	2
<i>Freie Werkstätten</i>	-0,1361	1

Tabelle 51: Ergebnisse des Verzerrungsfaktor-Modells bei Unterscheidung nach Art der Werkstatt

Tabelle 52 zeigt die Ergebnisse der Z-Statistik im Rahmen der o.g. Untersuchung. Die größten Abweichungen treten bei den freien Werkstätten hierbei im Jahr 2013 bei der Klasse $d_2 = 6$ mit einem Z-Wert von 2,52 auf. Im Jahr 2014 tritt die größte Deviation bei den Fünfen (Z-Wert: 2,04) auf.

	Klasse mit größter Deviation	Wert der Z-Statistik
2013		
<i>Markenwerkstätten</i>	$d_2 = 9$	1,58
<i>Freie Werkstätten</i>	$d_2 = 6$	2,52
2014		
<i>Markenwerkstätten</i>	$d_2 = 1$	2,25
<i>Freie Werkstätten</i>	$d_2 = 5$	2,04

Tabelle 52: Ergebnisse der Z-Statistik der zweiten Ziffern bei Unterscheidung nach Art der Werkstatt

Die Verteilung der zweiten Ziffern der freien Werkstätten in den Jahren 2013 und 2014 ist in Abbildung 24 dargestellt. Die entsprechenden Ergebnisse der Z-Statistik lassen sich hier sehr gut nachvollziehen: Im Jahr 2013 weisen die Sechsen und im Jahr 2014 die Fünfen und die Vieren die größten Abweichungen auf. Die Nullen sind wie bei dem Gesamtdatensatz in beiden Jahren unterrepräsentiert. Auf Basis dieser Verteilungen sind jedoch keine konkreten Abweichungsmuster zu erkennen, die z.B. auf gezielte Aufrundungen hinweisen.

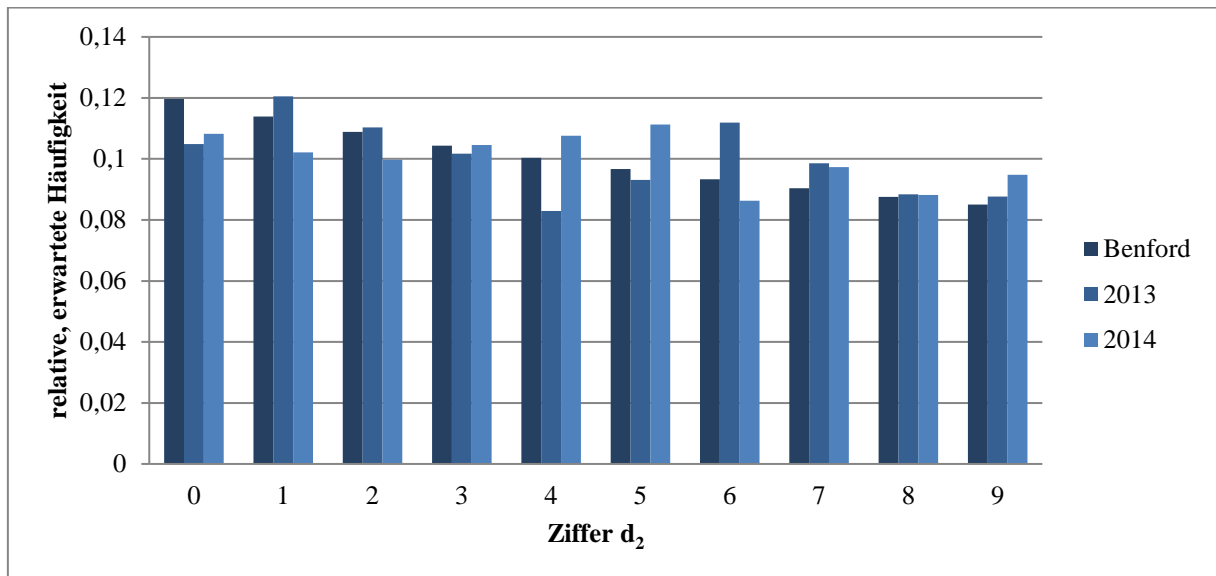


Abbildung 24: Verteilung der zweiten Ziffern der freien Werkstätten der Jahre 2013 und 2014

Im Rahmen eines weiteren Vergleichs der Datenqualität erfolgt die Separation des Datensatzes hinsichtlich der Automarke des Unfall-Kfz. Um Reputationskonflikte zu vermeiden, wurden die sieben analysierten Hersteller anonymisiert, sodass auf deren Namensnennung verzichtet wurde. Hierbei werden drei sog. „Premiumhersteller“, bezeichnet als Premiumhersteller A bis C, und vier sog. „Massenhersteller“, als Massenhersteller A bis D bezeichnet, betrachtet. Mit Ausnahme von Massenhersteller C, für welchen 2013 lediglich 117 und 2014 nur 194 Zahlen zur Verfügung stehen, liegen die Stichprobenumfänge der betrachteten Hersteller über dem Wert von $n = 300$. Zudem liegen die Stichprobenumfänge allesamt in einem ähnlichen Bereich, sodass die Ergebniswerte sehr gut vergleichbar sind. Im Falle weiterer Hersteller ist die Anzahl der zur Verfügung stehenden Daten zu gering, sodass ein Einbezug in die Untersuchung nicht möglich ist. Die Ergebnisse der Rangzahlen des Benford-Testes bei Applikation auf die zweiten Ziffern der Daten zeigt Tabelle 53. Die besten Resultate treten bei Premiumhersteller A und Massenhersteller D auf. Insbesondere sind die Ergebnisse über beide Jahre und bei einem Großteil der Testverfahren eindeutig und somit als sehr robust anzusehen. Die größten Deviationen treten bei Premiumhersteller C und Massenhersteller A auf. Beide befinden sich sowohl 2013 als auch 2014 jeweils unter den ersten drei Plätzen. Auch hier liefern alle Verfahren weitestgehend ähnliche Ergebnisse. Die exakten Zahlenwerte sind in Tabelle 88 in Anhang VII verzeichnet. Hier lässt sich erkennen, dass die Testgrößen von Premiumhersteller C und Massenhersteller A zwar höher ausfallen als die der übrigen Hersteller, die Unterschiede jedoch

nicht sehr deutlich in Erscheinung treten. Eine Aussage hinsichtlich Unterschieden in der Datenqualität zwischen Premiumherstellern und Massenherstellern lässt sich hingegen nicht treffen, da hier keine der Gruppen auffälligere Abweichungen aufweist.

Rangzahlen 2. Ziffern	n	χ^2	KS	Kui- per	$\frac{\chi^2}{n}$	d*	a*	Man- hattan	Tscheby- scheff	Ge- samt
2013										
<i>Premiumhersteller A</i>	350	7	7	7	7	7	6	7	7	7
<i>Premiumhersteller B</i>	383	1	3	4	3	3	3	4	1	2
<i>Premiumhersteller C</i>	442	4	2	1	4	4	1	3	4	3
<i>Massenhersteller A</i>	267	3	1	3	1	1	5	1	2	1
<i>Massenhersteller B</i>	406	5	6	5	5	5	7	5	5	6
<i>Massenhersteller C</i>	117	6	5	6	2	2	4	2	3	4
<i>Massenhersteller D</i>	838	2	4	2	6	6	2	6	6	5
2014										
<i>Premiumhersteller A</i>	491	6	7	5	6	6	7	7	7	7
<i>Premiumhersteller B</i>	520	4	4	2	5	5	4	5	4	4
<i>Premiumhersteller C</i>	542	2	1	1	3	2	2	3	1	2
<i>Massenhersteller A</i>	363	3	2	3	1	1	1	1	2	1
<i>Massenhersteller B</i>	519	1	5	4	2	3	3	2	5	3
<i>Massenhersteller C</i>	194	7	3	7	4	4	5	6	3	5
<i>Massenhersteller D</i>	1.096	5	6	6	7	7	6	4	6	6

Tabelle 53: Ergebnisse des Benford-Testes in Form von Rangzahlen bei Applikation auf die zweiten Ziffern bei Unterscheidung nach Kfz-Hersteller

Die Ergebnisse des Verzerrungsfaktor-Modells sind als Ergänzung der Vollständigkeit halber in Tabelle 54 angegeben. Hier liegen Premiumhersteller C und Massenhersteller A lediglich auf Plätzen im Mittelfeld. Aufgrund der in Abschnitt 4.3 dargelegten Problematik, sollte diesem Ergebnis jedoch im Grunde keine Bedeutung zugeschrieben werden, da das Verzerrungsfaktor-Modell aufgrund dessen Fehleranfälligkeit nur begrenzte Aussagekraft besitzt.

Ergebnisse Verzerrungsfaktor	VF	Rang
2013		
<i>Premiumhersteller A</i>	-0,1300	5
<i>Premiumhersteller B</i>	-0,1724	2
<i>Premiumhersteller C</i>	-0,1389	4
<i>Massenhersteller A</i>	-0,0610	7
<i>Massenhersteller B</i>	-0,1768	1
<i>Massenhersteller C</i>	-0,0806	6
<i>Massenhersteller D</i>	-0,1476	3
2014		
<i>Premiumhersteller A</i>	-0,0987	5
<i>Premiumhersteller B</i>	-0,1306	3
<i>Premiumhersteller C</i>	-0,0598	7
<i>Massenhersteller A</i>	-0,1288	4
<i>Massenhersteller B</i>	-0,1779	2
<i>Massenhersteller C</i>	-0,0851	6
<i>Massenhersteller D</i>	-0,1796	1

Tabelle 54: Ergebnisse des Verzerrungs-Faktormodells bei Unterscheidung nach Kfz-Hersteller

Um genauere Informationen über die Ursache der Abweichungen zu erhalten sind zusätzlich die Werte der Z-Statistik für Premiumhersteller C und Massenhersteller A zu betrachten, die in Tabelle 55 dargestellt sind. Die Klassen mit den jeweils größten Abweichungen sind in beiden Jahren und bei beiden Herstellern stets unterschiedlich, sodass hierbei keine konkreten Schlussfolgerungen getroffen werden können.

	Klasse mit größter Deviation	Wert der Z-Statistik
2013		
Premiumhersteller C	$d_2 = 8$	1,99
Massenhersteller A	$d_2 = 5$	2,21
2014		
Premiumhersteller C	$d_2 = 1$	2,87
Massenhersteller A	$d_2 = 9$	2,19

Tabelle 55: Ergebnisse der Z-Statistik der zweiten Ziffern von Premiumhersteller C und Massenhersteller A

Dies wird auch deutlich, wenn man sich die Verteilung der zweiten Ziffern in Abbildung 25 näher betrachtet. Diese sind zwar für beide Jahre und Hersteller unterschiedlich, zeigen jedoch in allen Fällen gravierende Abweichungen von der Benford-Verteilung. Gut zu erkennen sind hierbei jeweils die von der Z-Statistik entlarvten Ausreißer. Aufgrund der starken Diversität der Ergebnisse lassen sich jedoch auch hier keine Manipulationsmuster erkennen.

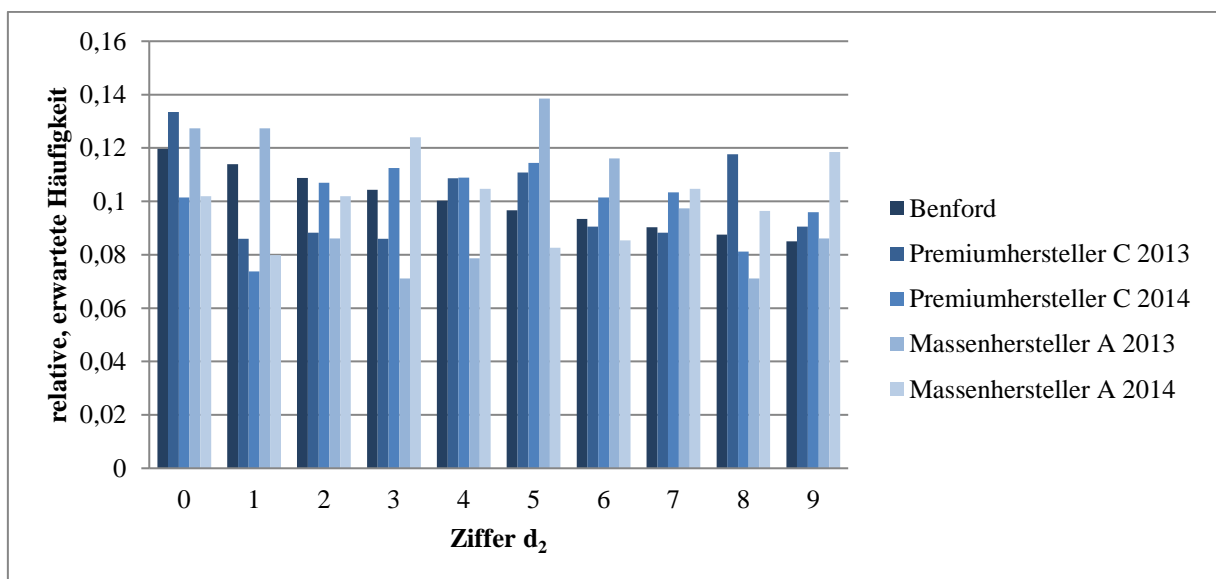


Abbildung 25: Verteilung der zweiten Ziffern von Premiumhersteller C und Massenhersteller A in den Jahren 2013 und 2014

Auf Basis dieser Resultate stellt sich die Frage, ob die Ursache der Abweichungen bei den Kfz-Herstellern genauer spezifiziert werden kann. Insbesondere ist von Interesse, ob die Abweichungen durch die freien Werkstätten oder durch die Vertragswerkstätten zu Stande kommen. Aus diesem Grund wird der nach Kfz-Herstellern separierte Datensatz weiter hinsichtlich der entsprechenden Art der Werkstatt aufgegliedert und der Benford-Test auf diese Daten angewendet. Hierzu sei angemerkt, dass diese Analyse im Falle von Massenhersteller A und C nicht

möglich ist, da bei weiterer Unterteilung die Anzahl der zur Verfügung stehenden Daten zu gering ist und die Mindestanforderung an den Stichprobenumfang nicht erfüllt wäre.

Die Ergebnisse der Kfz-Hersteller bei ausschließlich freien Werkstätten finden sich in Tabelle 56, die exakten Zahlenwerte sind Tabelle 89 in Anhang VII zu entnehmen. Hier liegen die Stichprobenumfänge zwar allesamt in einem ähnlichen Bereich, sodass die Ergebnisse sehr gut miteinander vergleichbar sind, jedoch wird der empfohlene Umfang von 300 Zahlen leicht unterschritten. Auffällig ist, dass v.a. die Zahlen von Premiumhersteller B hohe Abweichungen aufweisen. Die geringsten Deviationen treten wie oben bei Massenhersteller D auf. Ebenfalls von besonderer Relevanz ist das Ergebnis von Premiumhersteller C: Bei ausschließlicher Betrachtung der freien Werkstätten sind die Ergebnisse an und für sich nicht bzw. weit weniger auffällig. Im Jahr 2013 liegt Premiumhersteller C sogar lediglich auf dem vierten Platz. Dies lässt die Schlussfolgerung zu, dass die oben geschilderten hohen Abweichungen bei Premiumhersteller C v.a. auf die Vertragswerkstätten zurückzuführen sind. Dies wird im Folgenden genauer betrachtet.

Rangzahlen 2. Ziffern	n	χ^2	KS	Kui- per	$\frac{\chi^2}{n}$	d*	a*	Man- hattan	Tscheby- scheff	Ge- samt
2013										
<i>Premiumhersteller A</i>	117	4	4	3	3	2	5	2	3	3
<i>Premiumhersteller B</i>	162	1	1	2	2	3	4	3	1	2
<i>Premiumhersteller C</i>	108	5	3	5	4	4	1	4	4	4
<i>Massenhersteller B</i>	125	2	2	1	1	1	2	1	2	1
<i>Massenhersteller D</i>	262	3	5	4	5	5	3	5	5	5
2014										
<i>Premiumhersteller A</i>	141	4	3	4	4	4	3	4	4	4
<i>Premiumhersteller B</i>	170	1	1	1	1	1	1	2	1	1
<i>Premiumhersteller C</i>	117	3	2	5	2	2	2	1	2	2
<i>Massenhersteller B</i>	138	2	4	3	3	3	4	3	3	3
<i>Massenhersteller D</i>	360	5	5	2	5	5	5	5	5	5

Tabelle 56: Ergebnisse des Benford-Testes in Form von Rangzahlen bei Applikation auf die zweiten Ziffern der freien Werkstätten bei Unterscheidung nach Kfz-Hersteller

In Ergänzung sind in Tabelle 57 zur Vollständigkeit die Ergebnisse des Verzerrungsfaktor-Modells angegeben. Jedoch lässt sich auch hier erkennen, dass die Ergebnisse leicht von den Resultaten der übrigen Verfahren abweichen: Massenhersteller D liegt hier mit Abweichungen von 19,04% im Jahr 2013 und 13,63% in 2014 auf dem zweiten bzw. dritten Platz. Premiumhersteller B, der Hersteller mit den größten Abweichungen bei den freien Werkstätten, weist im Jahr 2013 lediglich den vierten Platz auf.

Ergebnisse Verzerrungsfaktor	VF	Rang
2013		
<i>Premiumhersteller A</i>	-0,1331	5
<i>Premiumhersteller B</i>	-0,1348	4
<i>Premiumhersteller C</i>	-0,1813	3
<i>Massenhersteller B</i>	-0,2130	1
<i>Massenhersteller D</i>	-0,1904	2
2014		
<i>Premiumhersteller A</i>	-0,1056	4
<i>Premiumhersteller B</i>	-0,1835	1
<i>Premiumhersteller C</i>	-0,1485	2
<i>Massenhersteller B</i>	-0,0874	5
<i>Massenhersteller D</i>	-0,1363	3

Tabelle 57: Ergebnisse des Verzerrungs-Faktormodells der freien Werkstätten bei Unterscheidung nach Kfz-Hersteller

Tabelle 58 zeigt die Ergebnisse der zweiten Ziffern der Aufgliederung nach Kfz-Herstellern bei ausschließlichem Einbezug der Markenwerkstätten. Die exakten Zahlenergebnisse sind in Tabelle 90 in Anhang VII dargestellt. Hierbei wird der empfohlene Mindeststichprobenumfang von $n = 300$ Zahlen weitestgehend eingehalten bzw. nur knapp unterschritten. Auch weichen die Umfänge nur gering voneinander ab, sodass auch hier eine hervorragende Komparabilität vorliegt. Die geschilderte Vermutung, dass die hohen Abweichungen bei Premiumhersteller C auf die Markenwerkstätten zurückzuführen sind, bestätigt sich hierbei eindeutig: In beiden Jahren liegt Premiumhersteller C klar auf dem ersten Platz. Auch zahlentechnisch liegen die Ergebnisse der Testgrößen zum Teil deutlich über denjenigen der übrigen Hersteller. Hinsichtlich

der Art der Hersteller lassen sich jedoch auch hier keine allgemeinen Schlussfolgerungen treffen.

Rangzahlen 2. Ziffern	n	χ^2	KS	Kui- per	$\frac{\chi^2}{n}$	d*	a*	Man- hattan	Tscheby- scheff	Ge- samt
2013										
<i>Premiumhersteller A</i>	233	3	5	4	2	2	5	2	2	3
<i>Premiumhersteller B</i>	221	5	3	5	4	3	3	5	2	4
<i>Premiumhersteller C</i>	334	2	1	2	1	1	2	1	1	1
<i>Massenhersteller B</i>	281	4	4	3	5	5	4	4	5	5
<i>Massenhersteller D</i>	576	1	2	1	3	4	1	3	4	2
2014										
<i>Premiumhersteller A</i>	350	4	3	3	4	4	4	4	3	4
<i>Premiumhersteller B</i>	350	3	4	4	2	2	3	3	2	3
<i>Premiumhersteller C</i>	425	1	1	1	1	1	1	1	1	1
<i>Massenhersteller B</i>	381	2	2	2	3	3	2	2	4	2
<i>Massenhersteller D</i>	736	5	5	5	5	5	5	5	5	5

Tabelle 58: Ergebnisse des Benford-Testes in Form von Rangzahlen bei Applikation auf die zweiten Ziffern der Markenwerkstätten bei Unterscheidung nach Kfz-Hersteller

Tabelle 59 stellt die Ergebnisse des Verzerrungsfaktormodells für die Markenwerkstätten dar. Auch hier liefert das Modell abweichende Resultate: So weicht der Mittelwert der Zahlen von Massenhersteller D im Jahr 2014 mit 19,05% am stärksten von dem zu erwartenden Benford-Mittelwert ab, obwohl die restlichen Verfahren die beste Benford-Annäherung signalisieren. Im Fall von Premiumhersteller C ist dies 2013 genau umgekehrt: Indizierten die anderen Verfahren die größten Wahrscheinlichkeiten für vorhandene Manipulationen, so weist das Verzerrungsfaktormodell die beste Benford-Konformität auf. Demzufolge scheint das Verzerrungsfaktormodell auch hier eher fragwürdige Ergebnisse zu liefern.

Ergebnisse Verzerrungsfaktor	VF	Rang
2013		
<i>Premiumhersteller A</i>	-0,1241	4
<i>Premiumhersteller B</i>	-0,1957	1
<i>Premiumhersteller C</i>	-0,1122	5
<i>Massenhersteller B</i>	-0,1675	2
<i>Massenhersteller D</i>	-0,1582	3
2014		
<i>Premiumhersteller A</i>	-0,0930	4
<i>Premiumhersteller B</i>	-0,1029	3
<i>Premiumhersteller C</i>	-0,1169	2
<i>Massenhersteller B</i>	-0,0497	5
<i>Massenhersteller D</i>	-0,1905	1

Tabelle 59: Ergebnisse des Verzerrungs-Faktormodells der Markenwerkstätten bei Unterscheidung nach Kfz-Hersteller

Die Ergebnisse der Z-Statistik in Tabelle 60 zeigen grundsätzlich – wie es auch zu erwarten ist – die gleichen Erkenntnisse wie bei der Analyse ohne Separation nach der Werkstattart: Die hohen Abweichungen der Vertragswerkstätten von Premiumhersteller C sind im Wesentlichen auf die Klassen der Achten (2013) und Einsen (2014) zurückzuführen, die mit 2,76 und 2,73 die mit Abstand höchsten Z-Werte liefern.

	Klasse mit größter Deviation	Wert der Z-Statistik
2013		
<i>Premiumhersteller C</i>	$d_2 = 8$	2,76
2014		
<i>Premiumhersteller C</i>	$d_2 = 1$	2,73

Tabelle 60: Ergebnisse der Z-Statistik für die zweiten Ziffern der Markenwerkstätten von Premiumhersteller C

Als weitere Teilbetrachtung wird in einem nächsten Schritt untersucht, ob es Unterschiede in der Datenqualität hinsichtlich der Art des Versicherungsfalls gibt. Hierzu wird der Datensatz der Partnerwerkstätten aufgeteilt in Kasko- und Haftpflichtschäden und der Benford-Test auf die zweiten Ziffern der enthaltenen Zahlen angewendet. Die Ergebnisse in Form von Rangzahlen finden sich in Tabelle 61. Der kleinste Umfang vorhandener Zahlen liegt bei $n = 1.116$, sodass der empfohlene Mindestwert von $n = 300$ deutlich überstiegen wird. Auch hier weichen die Umfänge nicht zu stark voneinander ab, dass eine gute Vergleichbarkeit der Ergebnisse gegeben ist.

Auffällig ist, dass beim Vergleich beider Jahre die Resultate nicht eindeutig sind. So weisen im Jahr 2013 die Haftpflichtschäden und im Jahr 2014 die Kaskoschäden eine höhere Abweichung auf, wobei die Unterschiede äußerst knapp ausfallen: Vergleicht man die individuellen Ergebnisgrößen in Tabelle 91 in Anhang VII so stellt man fest, dass Unterschiede zum Großteil erst in der zweiten bzw. dritten Nachkommastelle der Testgrößen erkennbar sind. Demzufolge lassen sich auf Basis dieser Ergebnisse keine eindeutigen Schlüsse ziehen. Fraglich bleibt, ob die Ergebnisse anders ausfallen, wenn man auch hierbei zwischen den freien Werkstätten und den Markenwerkstätten unterscheidet. Schließlich könnte es sein, dass bei bestimmten Arten von Schäden die Manipulationen schwerpunktmäßig in einer bestimmten Werkstattart vorgenommen werden.

Rangzahlen 2. Ziffern	n	χ^2	KS	Kui- per	$\frac{\chi^2}{n}$	d*	a*	Man- hattan	Tscheby- scheff	Ge- samt
2013										
<i>Kasko-Schaden</i>	2.712	1	2	1	2	2	2	2	2	2
<i>Haftpflicht-Schaden</i>	1.116	2	1	2	1	1	1	1	1	1
2014										
<i>Kasko-Schaden</i>	3.523	1	1	1	1	1	1	1	1	1
<i>Haftpflicht-Schaden</i>	1.585	2	2	2	1	2	2	1	2	2

Tabelle 61: Ergebnisse des Benford-Testes in Form von Rangzahlen bei Applikation auf die zweiten Ziffern bei Unterscheidung nach Art des Schadens

Auch Tabelle 62, welche die Ergebnisse des Verzerrungsfaktor-Modells bei Unterscheidung nach der Art des Schadens darlegt, zeigt, dass die Ergebnisse der Kasko- und Haftpflichtschäden nah beieinander liegen und sich nicht deutlich unterscheiden. Insgesamt weisen hier die Kasko-Schäden leicht höhere Abweichungen von dem Benford-Mittelwert auf.

Ergebnisse Verzerrungsfaktor	VF	Rang
2013		
<i>Kasko-Schaden</i>	-0,1415	1
<i>Haftpflicht-Schaden</i>	-0,1003	2
2014		
<i>Kasko-Schaden</i>	-0,0930	1
<i>Haftpflicht-Schaden</i>	-0,0903	2

Tabelle 62: Ergebnisse des Verzerrungs-Faktormodells bei Unterscheidung nach Art des Schadens

Die Ergebnisse in Form von Rangzahlen bei Anwendung des Benford-Testes auf die zweiten Ziffern mit Unterscheidung nach der Art des Schadensfalls der freien Werkstätten zeigt Tabelle 63. Der minimale Stichprobenumfang sowie eine äquivalente Größenordnung der Umfänge sind auch hier gegeben. Neue Erkenntnisse lassen sich aus der Untersuchung jedoch nicht gewinnen: Während im Jahr 2013 die Haftpflichtschäden erneut die größere Benford-Deviation aufweisen, fällt das Ergebnis im Jahr 2014 nicht eindeutig aus: Hier attestieren vier der Verfahren eine höhere Abweichung der Kasko-Schäden und vier bei den Haftpflicht-Schäden. Bei einem Blick auf die konkreten Zahlenwerte in Tabelle 92 in Anhang VII lässt sich ebenfalls feststellen, dass die Ergebnisgrößen der Kasko-Schäden und der Haftpflicht-Schäden nah beieinander liegen und demnach ebenfalls keine schlüssigen Aussagen zulassen.

Rangzahlen 2. Ziffern	n	X^2	KS	Kui- per	$\frac{X^2}{n}$	d*	a*	Man- hattan	Tscheby- scheff	Ge- samt
2013										
<i>Kasko-Schaden</i>	949	1	2	2	2	2	1	2	2	2
<i>Haftpflicht-Schaden</i>	329	2	1	1	1	1	2	1	1	1
2014										
<i>Kasko-Schaden</i>	1.166	1	1	1	2	2	1	2	2	1
<i>Haftpflicht-Schaden</i>	479	2	2	2	1	1	2	1	1	1

Tabelle 63: Ergebnisse des Benford-Testes in Form von Rangzahlen bei Applikation auf die zweiten Ziffern der freien Werkstätten bei Unterscheidung nach Art des Schadens

Die Ergebnisse des Verzerrungsfaktormodells sind in Tabelle 64 zusammengefasst. Wie bereits ohne die Separation nach freien Werkstätten und Vertragswerkstätten, ergibt sich sowohl im Jahr 2013 als auch 2014 eine höhere Abweichung bei den Kaskoschäden. Hier fällt die Differenz zwischen den Schadensarten jedoch höher aus.

Ergebnisse Verzerrungsfaktor	VF	Rang
2013		
<i>Kasko-Schaden</i>	-0,1743	1
<i>Haftpflicht-Schaden</i>	-0,1006	2
2014		
<i>Kasko-Schaden</i>	-0,1683	1
<i>Haftpflicht-Schaden</i>	-0,0555	2

Tabelle 64: Ergebnisse des Verzerrungs-Faktormodells der freien Werkstätten bei Unterscheidung nach Art des Schadens

Tabelle 65 fasst die Rangzahlen der Markenwerkstätten bei Unterscheidung nach der Schadensart zusammen. Hinsichtlich des Stichprobenumfanges ergeben sich keinerlei Einschränkungen der Untersuchung. Als Ergebnis zeigen sich auch hier widersprüchliche Ergebnisse, sodass aus der Untersuchung keine neuen plausiblen Erkenntnisse gewonnen werden können. In Tabelle 93 in Anhang VII sind die exakten Zahlenwerte der Testverfahren zusammengefasst. Die Werte der Testgrößen liegen hier ebenfalls sehr nah beieinander. Vergleicht man die Testgrößen der freien Werkstätten (Tabelle 92 in Anhang VII) und der Markenwerkstätten miteinander, so lässt

sich beobachten, dass die Abweichungen bei den freien Werkstätten grundsätzlich höher ausfallen als bei den Markenwerkstätten. Dies war jedoch bereits bei dem nicht weiter unterteilten Datensatz zu beobachten, sodass diese Beobachtung nur plausibel ist: Weisen freie Werkstätten im Allgemeinen eine höhere Benford-Deviation auf, so ist diese auch bei weiter unterteilten Datensätzen zu erwarten.

Rangzahlen 2. Ziffern	n	χ^2	KS	Kui- per	$\frac{\chi^2}{n}$	d*	a*	Man- hattan	Tscheby- scheff	Ge- samt
2013										
<i>Kasko-Schaden</i>	1.763	1	2	1	2	2	2	2	2	2
<i>Haftpflicht-Schaden</i>	787	2	1	2	1	1	1	1	1	1
2014										
<i>Kasko-Schaden</i>	2.357	1	1	1	2	2	1	1	2	1
<i>Haftpflicht-Schaden</i>	1.106	2	2	2	1	1	2	2	1	2

Tabelle 65: Ergebnisse des Benford-Testes in Form von Rangzahlen bei Applikation auf die zweiten Ziffern der Markenwerkstätten bei Unterscheidung nach Art des Schadens

Die Ergebnisse des Verzerrungsfaktor-Modells in Tabelle 66 indizieren hier ebenfalls eine höhere Abweichung von der Benford-Verteilung bei den Kasko-Schäden, was zumindest für 2013 im Vergleich zu den übrigen Verfahren einen Widerspruch impliziert.

Ergebnisse Verzerrungsfaktor	VF	Rang
2013		
<i>Kasko-Schaden</i>	-0,1233	1
<i>Haftpflicht-Schaden</i>	-0,0988	2
2014		
<i>Kasko-Schaden</i>	-0,1328	1
<i>Haftpflicht-Schaden</i>	-0,1044	2

Tabelle 66: Ergebnisse des Verzerrungs-Faktormodells der Markenwerkstätten bei Unterscheidung nach Art des Schadens

Die dritte und letzte Analyse des Datensatzes der Partnerwerkstätten bezieht sich auf das Erstzulassungsdatum der Unfall-Kfz. Diese Betrachtung soll Aufschluss darüber geben, ob das Alter des Kfz mit der Datenqualität in einem kausalen Zusammenhang steht. Schließlich könnte es sein, dass Halter im Falle eines Neuwagens aufgrund des höheren sachlichen und emotionalen Schadens eher dazu bereit sind, Falschangaben zu tätigen. Hierzu wurden die Schadensfälle hinsichtlich des Alters der Unfall-Kfz in sechs Gruppen unterteilt: Jünger als 2 Jahre, zwischen 2 und 4 Jahren, zwischen 5 und 7 Jahren, zwischen 8 und 10 Jahren, zwischen 11 und 13 Jahren sowie älter als 13 Jahre. Die Ergebnisse dieser Analyse in Form von Rangzahlen sind in Tabelle 67 dargestellt. Die exakten Zahlenwerte können Tabelle 94 in Anhang VII entnommen werden. Die minimal empfohlene Stichprobengröße wird hierbei in zwei von zwölf Fällen unterschritten, ansonsten unterscheiden sich die Umfänge maximal um den Faktor sieben, sodass eine Vergleichbarkeit der Ergebnisse noch gegeben ist.

Betrachtet man die Resultate muss man jedoch feststellen, dass sich der oben geschilderte Verdacht nicht bestätigt: Für das Jahr 2013 zeigen sich zwar für sehr alte und auch sehr neue Autos die höchsten Abweichungen, im Jahr 2014 rangieren diese Ergebnisgruppen jedoch in der Schlussgruppe auf den Plätzen fünf und sechs. Dies spricht grundsätzlich für wenig aufschlussreiche und plausible Ergebnisse. Paradoxerweise zeigen die einzelnen Testverfahren innerhalb der beiden Jahre jedoch relativ einheitliche Resultate, sodass diese zumindest innerhalb der Jahre als robust angesehen werden können. Betrachtet man die Zahlenwerte der einzelnen Verfahren, so erschließt sich auch ein möglicher Erklärungsgrund für die eigentlich widersprüchlichen Ergebnisse: Die Zahlen liegen insgesamt in einem relativ engen Wertebereich, sodass keine der Gruppen stark auffällige Ergebnisse aufweist. Dies hat im Umkehrschluss jedoch zur Folge, dass die relativen Ergebnisse zwischen den Jahren durchaus divergieren können. Dass die Werte innerhalb der Jahre robust sind, spricht im Grunde für die Effektivität der Testverfahren, da diese demnach die vorhandenen Abweichungen in Übereinstimmung ermitteln.

Rangzahlen 2. Ziffern	n	X^2	KS	Kui- per	$\frac{X^2}{n}$	d*	a*	Man- hattan	Tscheby- scheff	Ge- samt
2013										
< 2000	228	2	1	1	1	1	2	1	1	1
2000 – 2002	238	6	3	6	3	3	5	2	3	4
2003 – 2005	526	5	4	5	4	4	1	3	4	3
2006 – 2008	951	4	5	3	6	6	4	6	6	6
2009 - 2011	1.290	1	6	2	5	5	6	5	5	5
> 2012	502	3	2	4	2	2	3	4	2	2
2014										
< 2001	357	6	5	6	4	4	5	4	6	5
2001 – 2003	361	2	1	1	1	1	1	1	1	1
2004 – 2006	762	4	2	3	2	2	2	2	2	2
2007 – 2009	1.295	3	3	4	5	5	3	6	3	4
2010 - 2012	1.558	1	4	2	3	3	4	3	4	3
> 2013	629	5	6	5	6	6	6	5	5	6

Tabelle 67: Ergebnisse der zweiten Ziffern in Form von Rangzahlen des Datensatzes der Partnerwerkstätten bei Unterscheidung nach Datum der Erstzulassung des Unfall-Kfz

Tabelle 68 zeigt als Ergänzung zu den vorgestellten Ergebnissen der Testverfahren die ermittelten Größen des Verzerrungsfaktormodells, wobei auch diese keine eindeutigeren Ergebnisse liefern. Zwischen den beiden Jahren weisen alle Gruppen unterschiedliche Platzierungen auf, welche weiterhin von den obigen Ergebnissen abweichen. Vergleicht man die Ergebnisgrößen der Gruppen, so liegen lediglich die Schäden von Fahrzeugen mit einer Erstzulassung zwischen 2000 und 2002 im Jahr 2013 in einem deutlich erhöhten Bereich: Die dortige Abweichung liegt betragsmäßig mit 20,22% knapp sieben Prozentpunkte oberhalb der zweitplatzierten Gruppe. Alle anderen Werte liegen in einem sehr ähnlichen Bereich.

Ergebnisse Verzerrungsfaktor	VF	Rang
2013		
< 2000	-0,1341	2
2000 – 2002	-0,2022	1
2003 – 2005	-0,1283	4
2006 – 2008	-0,1185	6
2009 - 2011	-0,1301	3
> 2012	-0,1269	5
2014		
< 2001	-0,0991	6
2001 – 2003	-0,1019	5
2004 – 2006	-0,1467	2
2007 – 2009	-0,1330	3
2010 - 2012	-0,1287	4
> 2013	-0,1511	1

Tabelle 68: Ergebnisse des Verzerrungs-Faktormodells bei Unterscheidung nach Datum der Erstzulassung

6.3.3 Ergebnisse des Datensatzes des Expertisechecks

Im Falle des Expertisechecks erfolgt zunächst der Vergleich der Datenqualität der eingereichten Zahlen vor und nach der Prüfung durch einen Sachverständigen der Versicherung sowie des am Schluss tatsächlich ausgezahlten Rechnungsbetrages. Die Rangzahlen der statistischen Testverfahren für die zweiten Ziffern sind in Tabelle 69 dargestellt. Mit jeweils mehr als 17.000 Zahlen ist die Mindestanforderung an die Stichprobenumfänge mehr als überschritten. Zudem sind diese innerhalb der Jahre nahezu identisch, sodass ein Ergebnisvergleich ideal möglich ist.

Die geringsten Abweichungen treten im Falle der durch einen Gutachter geprüften Schadensfälle auf. Deutlich höher fallen die Deviationen bei den Zahlen vor dieser Prüfung sowie der am Ende tatsächlich ausgezahlten Beträge aus, wobei letztere die insgesamt größten Abweichungen aufweisen. Tabelle 95 in Anhang VIII gibt die Zahlenwerte der Testgrößen an. Hierbei

fällt auf, dass die Zahlen vor der Prüfung und die tatsächlich ausgezahlten Beträge in einer vergleichbaren Größenordnung liegen, wohingegen die Zahlen nach der Überprüfung durch einen Sachbearbeiter eine deutlich bessere Benford-Konformität aufweisen. Die Ergebnisse beider Jahre sind hierbei äquivalent.

Rangzahlen 2. Ziffern	n	χ^2	KS	Kui- per	$\frac{\chi^2}{n}$	d*	a*	Man- hattan	Tscheby- scheff	Ge- samt
2013										
<i>Vor der Prüfung</i>	17.128	2	2	2	2	1	2	2	1	2
<i>Nach der Prüfung</i>	17.127	3	3	3	3	3	3	3	3	3
<i>Tatsächliche Auszahlung</i>	17.125	1	1	1	1	2	1	1	2	1
2014										
<i>Vor der Prüfung</i>	21.085	2	1	1	2	2	2	1	2	2
<i>Nach der Prüfung</i>	21.083	3	3	3	3	3	3	3	3	3
<i>Tatsächliche Auszahlung</i>	21.077	1	2	2	1	1	1	2	1	1

Tabelle 69: Ergebnisse der zweiten Ziffern in Form von Rangzahlen des Datensatzes „Expertisecheck“ bei Unterscheidung nach Stand der Prüfung

Die Ergebnisse des Verzerrungsfaktor-Modells zeigt Tabelle 70. Wie bereits bei den vorherigen Betrachtungen divergieren diese jedoch grundsätzlich im Vergleich zu den Resultaten der übrigen Verfahren. Zum einen unterscheiden sich die Ergebnisse zwischen den Jahren, zum anderen stimmt lediglich der zweite Platz der Zahlen vor einer Prüfung durch einen Versicherungs-Sachbearbeiter im Jahr 2014 mit den oben ermittelten Platzierungen überein. In allen fünf anderen Fällen ergeben sich divergente Ergebnisse innerhalb des Rankings.

Ergebnisse Verzerrungsfaktor	VF	Rang
2013		
<i>Vor der Prüfung</i>	-0,1412	1
<i>Nach der Prüfung</i>	-0,1392	2
<i>Tatsächliche Auszahlung</i>	-0,1235	3
2014		
<i>Vor der Prüfung</i>	-0,1352	2
<i>Nach der Prüfung</i>	-0,1462	1
<i>Tatsächliche Auszahlung</i>	-0,1254	3

Tabelle 70: Ergebnisse des Verzerrungsfaktor-Modells des Datensatzes „Expertisecheck“ bei Unterscheidung nach Stand der Prüfung

Welche Häufigkeitsklassen jeweils für die Deviationen verantwortlich sind zeigen die Ergebnisse der Z-Statistik für die zweiten Ziffern in Tabelle 71. Zu erkennen ist, dass bei allen betrachteten Gruppen die Klasse der Ziffer null die größte Abweichung aufweist. Im Vergleich zu den anderen beiden Gruppen fällt diese bei den Zahlen nach der Prüfung durch einen Sachbearbeiter in beiden Jahren geringer aus.

	Klasse mit größter Deviation	Wert der Z-Statistik
2013		
<i>Vor der Prüfung</i>	$d_2 = 0$	5,12
<i>Nach der Prüfung</i>	$d_2 = 0$	1,42
<i>Tatsächlich ausgezahlter Betrag</i>	$d_2 = 0$	4,03
2014		
<i>Vor der Prüfung</i>	$d_2 = 0$	3,95
<i>Nach der Prüfung</i>	$d_2 = 0$	3,35
<i>Tatsächlich ausgezahlter Betrag</i>	$d_2 = 0$	5,03

Tabelle 71: Ergebnisse der Z-Statistik der zweiten Ziffern des Datensatzes des Expertisechecks

Selbiges müsste sich auch bei genauer Betrachtung der Ziffernverteilung zeigen. Die Verteilung der zweiten Ziffern des Datensatzes des Expertisechecks bei Unterscheidung nach dem Status der Prüfung im Jahr 2013 zeigt Abbildung 26. Die Testergebnisse lassen sich hier sehr gut wiedererkennen, da lediglich die Zahlen nach der Prüfung durch einen Sachverständigen eine

gute Annäherung an die Benford-Verteilung aufweisen. Bei den Zahlen vor der Prüfung sowie den tatsächlich ausgezahlten Beträgen hingegen ist die Deviation zum Teil deutlich erkennbar: In beiden Fällen tritt eine eindeutig zu geringe Anzahl an Nullen auf, während v.a. die Ziffern fünf bis neun leicht überrepräsentiert sind.

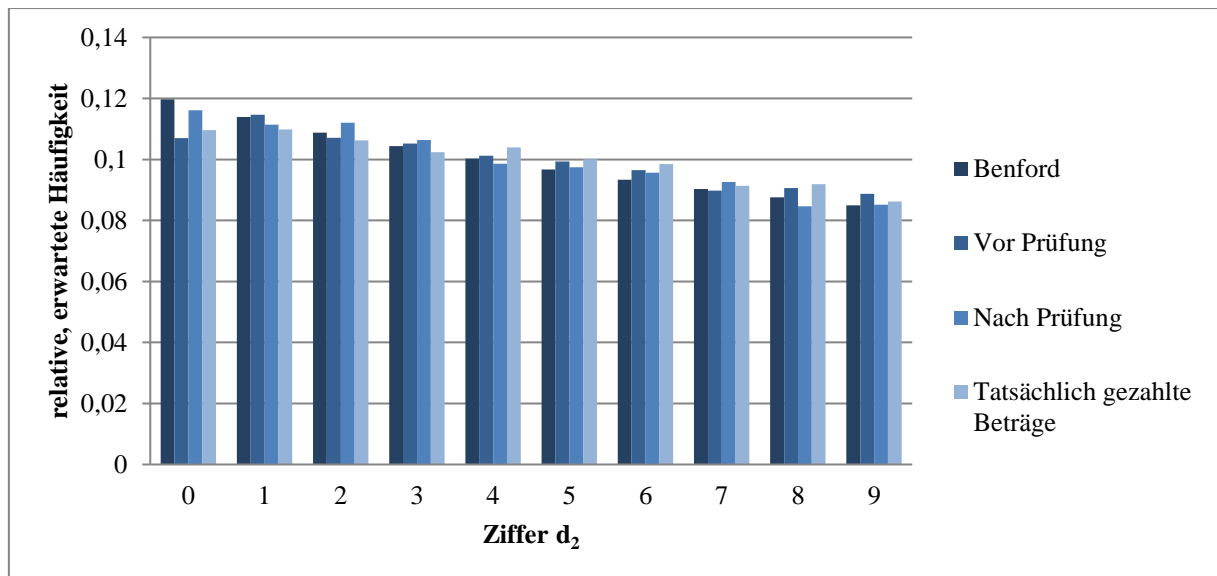


Abbildung 26: Verteilung der zweiten Ziffern des Datensatzes Expertisecheck des Jahres 2013 bei Unterscheidung nach Stand der Prüfung

Die Verteilung der Zahlen des Jahres 2014 ist in Abbildung 27 veranschaulicht. Hier zeigt sich grundsätzlich das gleiche Bild: Auch hier treten bei den Zahlen vor der Prüfung sowie den tatsächlich ausgezahlten Beträgen eine zu geringe Anzahl an Nullen auf. Die Zahlen nach der Prüfung zeigen eine leicht bessere Annäherung an die Benford-Verteilung.

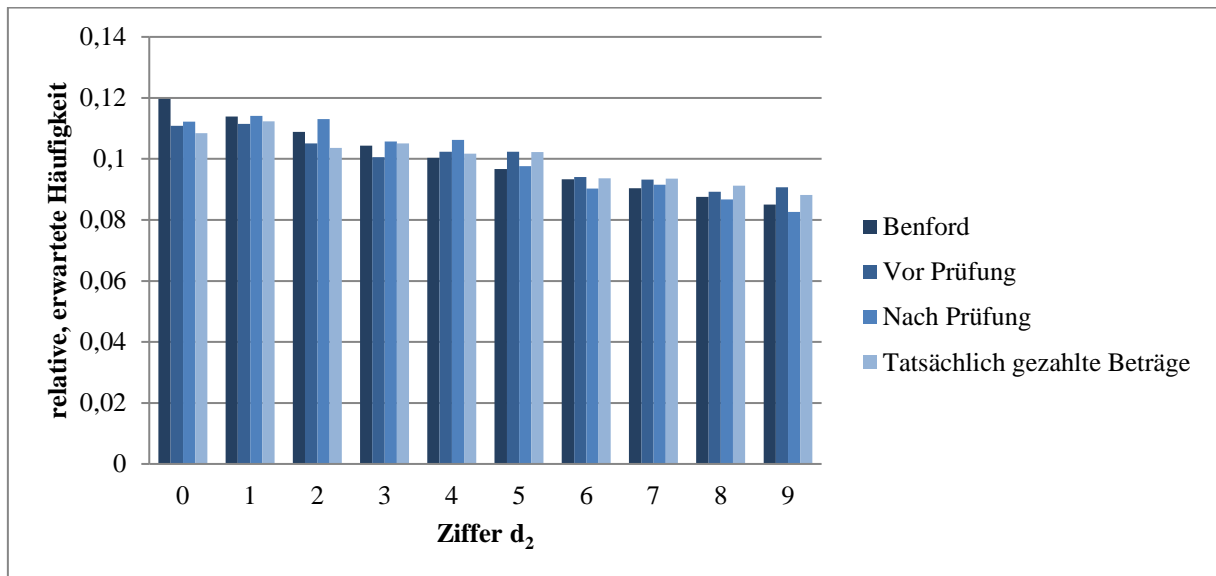


Abbildung 27: Verteilung der zweiten Ziffern des Datensatzes Expertisecheck des Jahres 2014 bei Unterscheidung nach Stand der Prüfung

In einem weiteren Schritt wird die Datenqualität hinsichtlich des Datums der Erstzulassung der Unfall-Kraftfahrzeuge analysiert. Wie auch schon bei der Untersuchung des Datensatzes der Partnerwerkstätten werden hierzu die Schäden hinsichtlich des Alters des Unfall-Kfz in die sechs Gruppen jünger als 2 Jahre, zwischen 2 und 4 Jahren, zwischen 5 und 7 Jahren, zwischen 8 und 10 Jahren, zwischen 11 und 13 Jahren sowie älter als 13 Jahren eingeteilt. Verwendet wurden hierbei die Beträge vor der Prüfung durch einen Sachverständigen, da diese unmittelbar durch den Versicherungsnehmer bzw. in Form eines Kostenvoranschlags durch die Werkstatt bei der Versicherung eingereicht worden sind und somit am ehesten Aussagen über das Manipulationsverhalten der Versicherungsnehmer bzw. der Werkstätten liefern. Die Ergebnisse als Rangzahlen sind in Tabelle 72 dargestellt, die detaillierten Ergebniswerte zeigt Tabelle 96 in Anhang VIII. Auch hier ergeben sich hinsichtlich der Stichprobenumfänge keinerlei Einschränkungen. Der Richtwert von $n = 300$ Zahlen ist auch hier deutlich übertroffen, zudem erstrecken sich die Umfänge von $n = 1.494$ bis $n = 5.676$ Zahlen und liegen somit in einem vergleichbaren Wertebereich.

Wie zuvor liefert die Analyse jedoch auch hier keine eindeutigen Ergebnisse: Die Testgrößen zeigen im Falle beider Jahre stark unterschiedliche Rangzahlen hinsichtlich der Abweichungen von der Benford-Verteilung. Betrachtet man die Daten der Fahrzeuge aus 2013, welche zwischen 2009 und 2011 zugelassen wurden, so ergibt sich im Durchschnitt die dritthöchste Abweichung von allen Untersuchungsgruppen. Unter den Einzelergebnissen der Testverfahren befinden sich jedoch ein sechster Platz, drei fünfte Plätze, ein dritter Platz, zwei zweite Plätze und

ein erster Platz. Dieses Beispiel zeigt, dass Testergebnisse hier eine nur geringe Robustheit besitzen. Dies wird untermauert, wenn man die Gesamtplatzierungen der beiden Jahre 2013 und 2014 vergleicht: Auch hier treten in beiden Jahren unterschiedliche Ergebnisse auf. Demzufolge liefert der Benford-Test bei der Unterscheidung nach dem Datum der Kfz-Erstzulassung keine eindeutigen Ergebnisse, was weitere Interpretationen und Schlussfolgerungen unmöglich macht.

Rangzahlen 2. Ziffern	n	χ^2	KS	Kui- per	$\frac{\chi^2}{n}$	d*	a*	Man- hattan	Tscheby- scheff	Ge- samt
2013										
< 2000	1.948	6	6	6	4	4	5	3	4	5
2000 – 2002	1.494	1	1	2	1	1	1	1	1	1
2003 – 2005	2.368	2	3	3	2	2	3	2	3	2
2006 – 2008	3.587	5	4	5	6	6	4	6	2	5
2009 - 2011	4.776	3	2	1	5	5	2	5	6	3
> 2012	2.094	4	5	4	3	3	6	4	5	4
2014										
< 2001	2.366	6	6	5	5	5	6	5	6	5
2001 – 2003	1.781	3	2	3	2	2	3	1	2	2
2004 – 2006	3.176	5	5	6	6	6	5	6	5	5
2007 – 2009	4.593	1	1	1	1	1	1	2	1	1
2010 - 2012	5.676	2	4	2	4	4	4	4	4	4
> 2013	2.457	4	3	4	3	3	2	3	3	3

Tabelle 72: Ergebnisse der zweiten Ziffern des Datensatzes des Expertisechecks bei Unterteilung nach Datum der Kfz-Erstzulassung

Der Vollständigkeit halber zeigt Tabelle 73 die Ergebnisse des Verzerrungsfaktormodells für die oben genannte Untersuchung. Zwar divergieren die Plätze des Rankings von denjenigen der übrigen Testverfahren, allerdings stimmen die Platzierungen in beiden Jahren fast überein.

Demnach liegen sowohl die neusten als auch die ältesten Fahrzeuge auf den letzten beiden Plätzen. Die größten Abweichungen treten bei der Gruppe der elf bis dreizehn-jährigen Kraftfahrzeugen auf.

Ergebnisse Verzerrungsfaktor	VF	Rang
2013		
< 2000	-0,1208	6
2000 – 2002	-0,1572	1
2003 – 2005	-0,1434	4
2006 – 2008	-0,1462	2
2009 - 2011	-0,1460	3
> 2012	-0,1290	5
2014		
< 2001	-0,1091	6
2001 – 2003	-0,1727	1
2004 – 2006	-0,1589	2
2007 – 2009	-0,1322	4
2010 - 2012	-0,1336	3
> 2013	-0,1261	5

Tabelle 73: Ergebnisse des Verzerrungsfaktor-Modells des Datensatzes „Expertisecheck“ bei Unterscheidung nach Datum der Erstzulassung des Unfall-Kfz

6.4 Bewertung der Ergebnisse

Insgesamt bleibt zunächst festzuhalten, dass der Benford-Test grundsätzlich zumindest eingeschränkt auf Kfz-Schadensfälle von Versicherungen angewendet werden kann, um Anhaltspunkte für systematische Manipulationen bei diesen zu erhalten. Aufgrund vorhandener Schwellenwerte, die durch Eigenbeteiligungen und Prämienerrhöhungen im Schadensfall entstehen, ist eine Anwendung des Benford-Testes auf die ersten Ziffern der Zahlen nicht möglich. Einer Applikation auf die zweiten Ziffern steht jedoch nichts entgegen.

Im Falle der Ergebnisbewertung wird grundsätzlich unterstellt, dass die mittels des Benford-Testes attestierten Abweichungen als Indikator für Manipulationen in den eingereichten Schadensfällen aufgefasst werden können. In diesem Kontext wird zeitgleich angenommen, dass mit steigendem Grad der Deviationen von der Benford-Verteilung die Wahrscheinlichkeit vorhandener fraudulenter Zahlen ebenfalls zunimmt. Eine Zusammenfassung der Ergebnisse der durchgeführten Untersuchung gibt Tabelle 74. Der Benford-Test selber liefert hierbei zwar keinerlei Informationen über die Ursache der Deviationen. Jedoch lassen sich die erhaltenen Ergebnisse sehr wohl durch sachlogische Schlussfolgerungen überprüfen und erklären.

Bei der Analyse des Datensatzes der Partnerwerkstätten im Falle der Separation nach der Art des Schadens sowie dem Datum der Erstzulassung der Unfall-Kfz konnten keine eindeutigen Ergebnisse gewonnen werden. Dies deutet darauf hin, dass anhand dieser beiden Merkmale keine Schlüsse auf eine erhöhte Wahrscheinlichkeit vorhandener Manipulationen gezogen werden können. Bei Unterscheidung nach der Art der Werkstatt und nach Hersteller des Kfz war dies jedoch sehr wohl der Fall.

Im Rahmen der Untersuchung der Zahlen der Partnerwerkstätten konnte gezeigt werden, dass die freien Werkstätten im Vergleich zu den Vertragswerkstätten eine insgesamt deutlich höhere Benford-Deviation aufweisen. Grundsätzlich deckt sich diese Erkenntnis sehr gut mit den zu erwartenden Resultaten: Vertragswerkstätten verfügen in der Regel über standardisierte Abrechnungs- und interne Qualitätssicherungssysteme, die Betrug durch beispielsweise zusätzliche eigentlich nicht benötigte Reparaturmaßnahmen oder Serviceleistungen deutlich erschweren. Beispielsweise sind sie im Rahmen der Schadensbehebung an die ausgegebenen Reparaturrichtlinien der Hersteller gebunden und dürfen zur Abrechnung nur die dort festgelegten Stundenumfänge heranziehen. Da solche bei freien Werkstätten nicht vorhanden sind und diese somit einen deutlich höheren Spielraum in der Abrechnung haben, liegt es nahe, dass hier vermehrt Manipulationen durchgeführt werden.²⁰⁴ Gleichzeitig werben freie Werkstätten häufig mit deutlich günstigeren Preisen für Ersatzteile sowie geringeren Stundensätzen als die Markenwerkstätten.²⁰⁵ Um dennoch profitabel wirtschaften zu können wäre es durchaus denkbar,

²⁰⁴ Vgl. Autobild.de, Werkstattrechnung – Das Kürzelchaos, <http://www.autobild.de/artikel/werkstattrechnung-38427.html> (Abgerufen am 02.02.2016), 2002.

²⁰⁵ Vgl. Handelsblatt-Online, Deutliche Preisunterschiede zwischen Markenwerkstatt und freiem Betrieb, <http://www.handelsblatt.com/auto/test-technik/autowerkstaetten-deutliche-preisunterschiede-zwischen-markenwerkstatt-und-freiem-betrieb/3052936.html> (Abgerufen am 29.01.2016), 2008.

dass die freien Werkstätten im Gegenzug zusätzliche nicht benötigte Reparaturen und höhere Zeitaufwände in ihr Preismodell fest einkalkulieren.

Bei Unterscheidung nach dem Hersteller der Unfall-Kfz zeigte sich v.a. bei Premiumhersteller C und Massenhersteller A eine erhöhte Benford-Deviation. Premiumhersteller A und Massenhersteller D lieferten bei der kumulierten Betrachtung beider Werkstatttypen die beste Annäherung an die Benford-Verteilung. Allgemeine Aussagen hinsichtlich der Datenqualität der Gruppen Premiumhersteller und Massenhersteller ließen sich nicht treffen. Bei genauerer Untersuchung traten die auffällig hohen Abweichungen bei Premiumhersteller C lediglich bei den Markenwerkstätten des Kfz-Herstellers auf, nicht jedoch bei den freien Werkstätten. Alle übrigen Hersteller lieferten hingegen unauffällige Ergebnisse bei den eigenen Vertragswerkstätten. Hieraus lässt sich schließen, dass Manipulationen von Schadensabrechnungen bei Vertragswerkstätten von Premiumhersteller C mit einer deutlich höheren Wahrscheinlichkeit einhergehen, als bei den restlichen Herstellern. Eine mögliche Erklärung für die unterschiedliche Datenqualität speziell bei den Markenwerkstätten könnte die divergente Effektivität eigener interner Qualitätssicherungs- bzw. Betrugserkennungssysteme der einzelnen Hersteller sein. Demzufolge wäre es möglich, dass die anderen Hersteller im Vergleich zu Premiumhersteller C Systeme verwenden, die den vorsätzlichen Betrug besser aufspüren oder aber artifizielle Erhöhungen der Rechnungsbeträge zuverlässiger erkennen und anzeigen.

Auch die aus der Analyse des Datensatzes des Expertisechecks gewonnenen Ergebnisse erwiesen sich im Nachhinein als durchaus plausibel. Verglichen wurden hierbei zunächst die Rechnungsbeträge der Kostenvorschläge vor und nach der Prüfung durch einen Sachverständigen der VKB. Hierbei zeigten die Zahlen nach der Prüfung eine deutlich bessere Konformität mit der Benford-Verteilung auf. Dies war auch in der Form zu erwarten, da zu hohe Beträge genauestens geprüft werden und damit z.B. zusätzlich durchgeführte Maßnahmen, die eigentlich in keinem Zusammenhang mit dem Unfallhergang stehen, oder zu hoch angesetzte Reparaturpauschalen erkannt und herausgestrichen werden können. Somit liegt es nahe, dass diese Zahlen mit einer sehr geringen Wahrscheinlichkeit vorhandener Manipulationen einhergehen. Die größten Deviationen traten bei den tatsächlich an die Versicherten ausgezahlten Beträgen auf, was auf dem ersten Blick zunächst abwegig erscheint. Doch auch für dieses Ergebnis existiert eine nachvollziehbare Erklärung: Die für die Bearbeitung des jeweiligen Schadensfalls zuständigen Sachbearbeiter können im Rahmen ihrer Arbeitsrichtlinien in begründeten Fällen von den Prüfgutachten abweichen und Zu- oder Abschläge auf die auszahlenden Beträge vornehmen.

Hierf8r k8nnen mehrere Ursachen in Frage kommen: Zum Beispiel w8re es eine M8glichkeit, dass dem Pr8fungutachter nicht alle Informationen zur Verf8gung stehen, um den Schadensfall vollst8ndig zu bearbeiten. So k8nnen als Erg8nzung beispielsweise im Nachgang Fotos vom Unfall-Kfz vorgelegt werden, die eine ver8nderte Beurteilung des Schadensfalls n8tig machen. Auch werden in einigen seltenen F8llen Gutachten wiederholt durch Sachverst8ndige durchgef8hrt, was zu Abweichungen zwischen den eingereichten Rechnungen und den tats8chlich ausgezahlten Rechnungsbetr8gen f8hrt. Diese durch systematische Eingriffe entstandenen Divergenzen k8nnen im Umkehrschluss auch Verzerrungen in den Ziffernverteilungen zur Folge haben und somit zu Abweichungen von dem Benfordschen Gesetz f8hren.²⁰⁶ Dies zeigt gleichzeitig, dass nicht nur mutwillig vorgenommene Manipulationen zu Abweichungen von der Benford-Verteilung f8hren, sondern diese auch Folge eines systematisch selektiven Arbeitsablaufes sein k8nnen.

Eine weitere Analyse der Daten des Expertisechecks bei Unterscheidung nach dem Datum der Erstzulassung lieferte wie bereits bei dem Datensatz der Partnerwerkst8tten keine eindeutigen Ergebnisse. Dies deutet darauf hin, dass das Alter des Autos kein Merkmal darstellt, dass mit einer erh8hten bzw. geringeren Manipulationswahrscheinlichkeit korreliert.

Untersuchungsgruppe	H8chste Benford-Abweichung	Niedrigste Benford-Abweichung
Partnerwerkst8tten		
<i>Art der Werkstatt</i>	Freie Werkst8tten	Markenwerkst8tten
<i>Kfz-Hersteller</i>	Premiumhersteller C, Massenhersteller A	Premiumhersteller A, Massenhersteller D
<i>Art des Schadens</i>	Keine Aussage m8glich	
<i>Datum der Erstzulassung</i>	Keine Aussage m8glich	
Expertisecheck		
<i>Zahlen vor/nach Pr8fung</i>	Tats8chlich gezahlter Betrag	Nach Pr8fung
<i>Datum der Erstzulassung</i>	Keine Aussage m8glich	

Tabelle 74: Zusammenfassung der Ergebnisse der Anwendung des Benford-Tests auf die zweiten Ziffern der Kfz-Schadensf8lle

²⁰⁶ Schriftliche Auskunft durch Herrn Dr. Sandro Hartmann, Inverso GmbH vom 28.10.2015.

Auf Basis der erhaltenen Ergebnisse erscheint die Annahme begründet, dass eine erhöhte Benford-Deviation in Daten der Kfz-Unfallsschäden auf einen erhöhten Anteil fraudulenter Zahlen zurückzuführen ist. Die gewonnenen Erkenntnisse konnten faktisch untermauert werden, so dass die Resultate aus 6.3 als plausibel angesehen werden können. Zwar lassen sich die Manipulationen mittels des Benford-Testes nicht sicher beweisen, jedoch können die Ergebnisse als ein wichtiger Indikator für solche interpretiert werden.

Auch wenn es für die Versicherungen aufgrund der zu geringen Anzahl an Zahlen nicht möglich ist, den Benford-Test auf Einzelfälle anzuwenden, ist es durchaus empfehlenswert diesen als Instrumentarium zur Betrugserkennung zu implementieren. So können mittels der gewonnenen Erkenntnisse Risikogruppen von Schadensfällen identifiziert werden, die mit einer besonders hohen Wahrscheinlichkeit vorhandener Manipulationen einhergehen. Diese sollten dann systematischen tiefergehenden Analysen unterzogen werden. Im Umkehrschluss kann bei Fällen mit geringerer Manipulationswahrscheinlichkeit auch der Examinierungsprozess eingeschränkt und Ressourcen somit effizienter allokiert werden. Auch sollten die bestehenden Strukturen kritisch überprüft und interne Qualitätssicherungssysteme in Kooperation mit den Partnerwerkstätten weiterentwickelt werden, um die Durchführbarkeit von Manipulationen einzuschränken.

7. Zusammenfassung und Ausblick

Im Rahmen dieser Forschungsarbeit wurde der Benford-Test auf seine Effektivität und Einschlägigkeit zur Erkennung von systematischem Betrug und Manipulationen getestet. Als Forschungsansatz wurde hierbei die Methodik gewählt, den Benford-Test auf Datensätze anzuwenden, in welchen Manipulationen sicher bzw. mit sehr großer Wahrscheinlichkeit vorhanden sind. Wenn der Benford-Test dazu in der Lage ist, diese Deviationen aufzudecken bzw. erkannte Abweichungen von der Benford-Verteilung plausibel erklärt werden können, ist von einer Eignung des Tests in dem untersuchten Anwendungsfeld auszugehen. Appliziert wurde der Benford-Test hierbei im Rahmen von zwei empirischen Untersuchungen auf historische volkswirtschaftliche Zahlen Deutschlands sowie auf Kfz-Schadensfälle einer Versicherungsgesellschaft.

Im Vorfeld dieser Untersuchungen wurde zunächst geprüft, wie eine auftretende Abweichung von der Benford-Verteilung idealerweise ermittelt werden kann, welche der bekannten statistischen Testverfahren und Maßzahlen sich also besonders für eine Anwendung im Rahmen des Benford-Testes eignen. Hierzu wurden in einem ersten Schritt mittels der Monte-Carlo-Simulation Benford-verteilte Datensätze mit verschiedenen Stichprobenumfängen generiert. In diesen wurden im Anschluss nach dem Zufallsprinzip Zahlen durch gleichverteilte Zahlen substituiert und so vorhandene Manipulationen in den Datensätzen simuliert. Daraufhin wurden die Benford-Deviationen der ersten, zweiten und ersten beiden Ziffern in Kombination mittels unterschiedlicher statistischer Anpassungstests und Maßzahlen ermittelt und auf Basis dieser Ergebnisse deren Eignung zur Aufspürung von fraudulentem Datenmaterial überprüft. Insbesondere war hierbei von Interesse, in welchem Bereich von Stichprobengrößen die Verfahren die zuverlässigsten Ergebnisse liefern bzw. ob bei diesen ein stichprobeninvariantes Verhalten vorliegt. Letzteres wäre genau dann der Fall, wenn die Testverfahren bei gleichen relativen Anteilen getürkter Zahlen für sämtliche Stichprobenumfänge äquivalente Zahlenwerte hervorbringen. Als Ergebnis konnte hierbei zusammenfassend gezeigt werden, dass sämtliche Verfahren zunächst einmal bei einer Anwendung auf die ersten und zweiten Ziffern einer Zahl im Rahmen des Benford-Testes angewendet werden können. Bei Applikation des Benford-Testes auf die ersten beiden Ziffern in Kombination sind der Chi²-Anpassungstest und der Kuiper-Test aufgrund unzuverlässiger Ergebnisse nur eingeschränkt geeignet. Ein Einsatz des Verfahrens der Z-Statistik empfiehlt sich vorwiegend, um bei bereits aufgespürten Abweichungen diejenigen Ziffern mit der größten Deviation zu identifizieren. So können nähere Informationen bzgl. der

Ursachen der Abweichungen gewonnen werden. Da ein Einsatz des von *Nigrini* entwickelten Verzerrungsfaktor-Modells zum Teil zu widersprüchlichen Ergebnissen führte, sollte auf die Anwendung dieses Verfahrens im Rahmen des Benford-Testes verzichtet werden. Hinsichtlich des Stichprobenumfanges lässt sich im Allgemeinen festhalten, dass die zu untersuchenden Datensätze für aussagekräftige Ergebnisse mindestens ca. 300 Werte beinhalten sollten. Ein stichprobeninvariantes Verhalten konnte bei dem KS-Test sowie bei dem Abstand über die arithmetischen Mittelwerte (a^*) beobachtet werden, sodass deren Ergebnisse uneingeschränkt vergleichbar sind. In hohem Maße stichprobenvariant sind der χ^2 -Anpassungstest sowie die Z-Statistik. Hier sind Vergleiche zwischen den Ergebniswerten nur bei übereinstimmenden Stichprobenumfängen möglich. Alle anderen Verfahren weisen eine eingeschränkte Stichprobeninvarianz auf. Dies bedeutet nichts anderes, als dass Ergebniswerte vergleichbar sind, sofern sich die Stichprobenumfänge lediglich um den Faktor fünf bis zehn unterscheiden. Weiterhin konnte gezeigt werden, dass aufgrund des Gesetzes der Großen Zahlen alle Signifikanztests bei großen Stichprobenumfängen die Benford-Verteilung bereits bei kleinsten Deviationen ablehnen. Aus diesem Grund ist es im Rahmen des Benford-Testes geeigneter, auf Interpretationen hinsichtlich der statistischen Signifikanz vollständig zu verzichten. Stattdessen sollten die Ergebnisgrößen lediglich für eine vergleichende Analyse eingesetzt werden. Zu diesem Zweck erweist es sich als empfehlenswert, zwischen den zu vergleichenden Untersuchungsobjekten Rangzahlen hinsichtlich der relativen Abweichung von der Benford-Verteilung zu ermitteln. Die erstplatzierten Objekte weisen hierbei die größten Benford-Abweichungen und damit den größten Anteil vorhandener Manipulationen auf. Angewendet werden sollte der Benford-Test auf die ersten und zweiten Ziffern einer Zahl. Die Applikation auf die ersten beiden Ziffern in Kombination liefert hingegen nur einen geringen Mehrwert, da die Manipulationserkennung hier latent unzuverlässiger ist.

Die erste empirische Untersuchung im Rahmen der Wirksamkeitsprüfung des Benford-Testes widmete sich der Analyse historischer volkswirtschaftlicher Zahlen Deutschlands in der Zeit von 1918 bis 1990. In diesem Kontext wurde insbesondere die Datenqualität während der verschiedenen existierenden Staatsformen in diesem Zeitfenster miteinander verglichen. Namentlich waren dies die Weimarer Republik (1918 – 1932), das NS-Regime (1933-1945), die Deutsche Demokratische Republik (1955 – 1990) und die Bundesrepublik Deutschland (1952 – 1990). Letztere sollte hierbei die beste Benford-Konformität aufweisen, da die BRD als gestandene parlamentarische Demokratie mit einer äußerst geringen Anzahl durchlebter ökonomi-

scher Krisen und Kriegsbeteiligungen nur wenig Anhaltspunkte für systematische Manipulationen eigener Wirtschaftsdaten mit sich bringt. Die größten Abweichungen würde man sicherlich unter den diktatorisch geführten Regierungen wie den Nationalsozialisten erwarten, da diese mit einem systematisch geplanten zweiten Weltkrieg allen Grund dazu gehabt hätten, eigene Wirtschaftsdaten zu fälschen, um Maßnahmen zur Kriegsvorbereitung zu verheimlichen oder aber die eigene Leistungsfähigkeit zu beschönigen. In der Folge müsste der Benford-Test bei entsprechender Eignung dazu in der Lage sein, solche möglichen Manipulationen zu erkennen und anzuzeigen. Um dieser Fragestellung nachzugehen wurden die Häufigkeiten der ersten und zweiten Ziffern der Datensätze ausgezählt und die Benford-Abweichung mit dem χ^2 -Anpassungstest, dem KS-Test, dem Kuiper-Test sowie den Maßzahlen χ^2/n , d^* , a^* , der Manhattan- und der Tschebyscheff-Distanz bestimmt. Die Ergebnisse des Verzerrungsfaktor-Modells wurden nicht mit in die Auswertung einbezogen. Mittels der Z-Statistik wurden im Falle auffälliger Deviationen die Ziffern mit den größten Abweichungen ermittelt. Im Zuge dessen konnte zunächst gezeigt werden, dass der gesamte Datensatz sowohl für die ersten als auch für die zweiten Ziffern eine nahezu ideale Benford-Verteilung aufwies. In einem zweiten Schritt wurde eine individuelle Auswertung für die unterschiedlichen Regierungsformen durchgeführt. Zu diesem Zweck wurden auf Basis der genannten Testverfahren und Maßzahlen Rangziffern zwischen den Staatsformen auf Basis der Benford-Abweichung ermittelt und diese so in eine effiziente Prüfreihefolge gebracht. Hierbei zeigte sich bei der Bundesrepublik Deutschland die beste Benford-Konformität. Die größten Deviationen wies das Nationalsozialistische Regime, gefolgt von der Weimarer Republik auf. Die Daten der Deutschen Demokratischen Republik lagen insgesamt auf dem dritten Platz. Im Rahmen einer Einzeljahresauswertung konnten zudem weitere wichtige Erkenntnisse gewonnen werden: Im Falle der Weimarer Republik und des NS-Regimes wurden einzelne Jahre identifiziert, die auffällig hohe Benford-Deviationen zeigten und somit ursächlich für die hohe Gesamtabweichung dieser Staatsformen sind. Im Falle des NS-Regimes waren dies die Jahre 1936, 1939 und 1941/42. Das Jahr 1936 war hierbei in erster Linie durch einen vehementen Ausbau der Wehrmacht sowie weitere Vorbereitungsmaßnahmen auf den zweiten Weltkrieg geprägt. Da die gesamten Kriegsvorbereitungen äußerst kostspielig waren, das Regime aber auf der anderen Seite auf die Unterstützung der Bevölkerung angewiesen war, liegt hier durchaus die Vermutung nahe, dass eigene Wirtschaftszahlen nach oben korrigiert wurden, um die eigene Ausgangslage zu beschönigen. Die Jahre 1939 und 1941/42 lagen beide innerhalb der Zeitspanne des zweiten Weltkriegs und lieferten die einzigen

veröffentlichten Zahlen dieser Zeit. Hier könnte der gezielte Täuschungsversuch der Kriegsgegner ursächlich für eine Benford-Deviation sein, um diese über eine vermeintlich bessere wirtschaftliche oder militärische Lage zu täuschen. In der Weimarer Republik lieferten die Jahre 1919, 1924 und 1931 auffällig hohe Abweichungen von der Benford-Verteilung, wobei im Ansatz ein Muster für systematische Aufrundungen erkennbar war. Alle drei auffälligen Jahre lagen befanden sich unmittelbar im Kontext historischer Krisen. 1919 erfolgte mit dem Versailler Vertrag ein umfassendes Schuldgeständnis Deutschlands zum Ausbruch des ersten Weltkrieges, was mit dramatischen ökonomischen Sanktionen und Konsequenzen einherging. Das Jahr 1924 lag in der Hochphase der deutschen Hyperinflation, welche zum einen mit einer schlechten volkswirtschaftlichen Ausgangslage einherging und zum anderen die korrekte Festsetzung von Preisen zur Leistungsbewertung in der Praxis unmöglich machte. Das Jahr 1931 war gekennzeichnet durch eine der verheerendsten Weltwirtschaftskrisen der menschlichen Historie, welcher die damaligen Deutschen Politiker mit einer im Nachhinein als katastrophal zu bewertenden Deflationspolitik versuchten zu entgegnen. Somit liegt es auch hier nahe, dass die Regierung Wirtschaftsdaten nach oben anpasste, um die getroffenen Krisenentscheidungen als gewinnbringend zu verkaufen. Bei Einzeljahresbetrachtung der DDR und der BRD konnten keine auffälligen Jahre robust identifiziert werden, jedoch überstiegen in 70% der betrachteten Jahre die Abweichungen der DDR diejenigen der BRD. Demnach kann davon ausgegangen werden, dass in der Deutschen Demokratischen Republik über nahezu den gesamten Zeitraum Manipulationen in geringen Umfängen durchgeführt worden sind. Auch dies stellte sich als nicht verwunderlich heraus, da die Regierung der DDR über den gesamten Zeitraum zum einen mit einer desaströsen wirtschaftlichen Lage umgehen und des Weiteren ein umfassendes Sozialsystem quersubventionieren musste. Gleichzeitig stand sie mit der Bundesrepublik Deutschland im stetigen Wettbewerb um die Fragestellung, welches der politischen Systeme das leistungsfähigere ist. Hier liegt es auf der Hand, dass die Kommunisten die offensichtlich schlechtere Ausgangslage lange Zeit versuchten, gezielt beschönigt darzustellen. Dass die Bundesrepublik Deutschland die beste Benford-Konformität liefert, zeigte sich ebenfalls als durchweg nachvollziehbar und plausibel. Aufgrund der exzellenten wirtschaftlichen Entwicklung über viele Jahre hinweg sowie der bis dato nicht vorhandenen Involvierung in kriegerische Auseinandersetzungen wären Abweichungen von der Benford-Verteilung im Grunde nicht zu begründen gewesen. Dieses Teilresultat kann somit als „Kontrollergebnis“ hinsichtlich der Wirksamkeit des Benford-Testes aufgefasst werden. Demzufolge zeigen sich alle Resultate im Rahmen der Anwendung auf volkswirtschaftliche Daten als durchweg plausibel. Wie bereits in

vorherigen Untersuchungen gezeigt wurde, spricht dies somit grundsätzlich dafür, dass der Benford-Test im Rahmen der forensischen Betrugsaufdeckung von Wirtschaftsdaten effektiv und effizient erste Anzeichen für Manipulationen aufspüren kann.

Die zweite empirische Anwendung des Benford-Testes erfolgte auf Versicherungsdaten in Form von Kfz-Unfallschäden. Hierzu wurden Datensätze der Jahre 2013 und 2014, die von der Inverso GmbH in Kooperation mit der Versicherungskammer Bayern zur Verfügung gestellt wurden, näher untersucht. Aufgrund zu geringer Anzahlen an Daten für die einzelnen Schadensfälle war eine individuelle Anwendung des Benford-Testes auf diese nicht möglich. Vielmehr wurde in diesem Kontext die Benford-Deviation in Abhängigkeit verschiedener Eigenschaften der Schadensfälle bestimmt und diejenigen Ausprägungen ermittelt die mit der größten Wahrscheinlichkeit vorhandener Manipulationen einhergehen. Auch hier mussten sämtliche durch den Benford-Test ermittelten Abweichungen durch plausibel darzulegende Argumente fundiert werden können, sofern sich der Benford-Test auf Kfz-Unfallschäden applizieren lässt. Um diese These zu bestätigen oder zu widerlegen wurden die Ziffernhäufigkeiten für die ersten und zweiten Ziffern der zur Verfügung stehenden Daten ausgezählt und die Benford-Abweichung ermittelt. Zum Einsatz kamen hierbei die gleichen Verfahren wie bei der ersten empirischen Untersuchung. Die Z-Statistik sowie das Verzerrungsfaktormodell kamen auch hier lediglich ergänzend zur Anwendung und wurden nicht mit in die Hauptuntersuchung einbezogen. Zur Verfügung standen insgesamt zwei Teildatensätze: Ein Datensatz, welcher sich aus Daten der Partnerwerkstätten der VKB zusammensetzte sowie einem Datensatz, welcher ausschließlich Schadensfälle beinhaltet, die Gegenstand eines Prüfgutachtens durch einen Sachverständigen waren. Zunächst konnte hierbei gezeigt werden, dass die aggregierten Zahlen keine Benford-Verteilung für die ersten Ziffern aufweisen. Grund hierfür sind mit großer Wahrscheinlichkeit Selbstbeteiligungen der Versicherungsnehmer oder aber ggf. drohende Erhöhungen der zu entrichtenden Versicherungsprämien im Schadensfall. Diese führen dazu, dass Schäden erst ab Erreichen einer bestimmten Schadenshöhe bei der Versicherung gemeldet werden und wirken so wie vorhandene Schwellenwerte in den Datensätzen. Demzufolge musste von einer Analyse der ersten Ziffern im Fall von Kfz-Schadensfällen abgesehen werden. Für die zweiten Ziffern zeigte sich hingegen insgesamt eine sehr gute Benford-Verteilung, sodass einer Applikation des Benford-Testes auf diese nichts entgegen stand. Im Anschluss wurden die Datensätze hinsichtlich bestimmter Eigenschaften weiter untergliedert und der Benford-Test auf die verschiedenen Ausprägungen angewendet. Im Falle der Partnerwerkstätten waren dies die Attri-

bute der *Werkstattart* (Markenwerkstatt, freie Werkstatt), der *Kfz-Hersteller*, der *Art des Schadens* (Kasko-Schaden, Haftpflichtschaden) und das *Datum der Erstzulassung* des Unfall-Kfz. Der Datensatz des Expertisechecks wurde hinsichtlich der Attribute nach dem *Stand der Prüfung* im Gutachtenprozess und dem *Datum der Erstzulassung* des Unfall-Kfz unterschieden. Bei Separation nach der Art der Werkstatt, in welcher die Schäden repariert wurden, konnte gezeigt werden, dass die eindeutig höheren Abweichungen bei den freien Werkstätten zu beobachten sind. Da diese im Gegensatz zu den Markenwerkstätten nicht über standardisierte Abrechnungs- bzw. Betrugserkennungssysteme verfügen, ist dies zunächst einmal plausibel. Somit ist ein Betrug durch die freien Werkstätten deutlich leichter zu realisieren und somit auch als wahrscheinlicher anzusehen. Ein weiterer Grund könnten auch die deutlich günstigeren Preise der freien Werkstätten sein, bei welchen im Grunde nicht benötigte Service- oder Reparaturmaßnahmen in die Preisberechnungsmodelle einkalkuliert werden. Unter den verschiedenen Kfz-Herstellern wiesen Premiumhersteller C und Massenhersteller A die größten Wahrscheinlichkeiten für vorhandene Manipulationen auf. Premiumhersteller A und Massenhersteller D zeigten die beste Benford-Konformität. Im Falle von Premiumhersteller C konnte zudem gezeigt werden, dass Abweichungen in erster Linie bei den Markenwerkstätten zu beobachten sind. Hierbei liegt die Vermutung nahe, dass die internen Betrugserkennungssysteme bei Premiumhersteller C im Vergleich mit den übrigen Herstellern weniger effizient funktionieren. Die Unterscheidung nach der Schadensart sowie dem Datum der Erstzulassung des Unfall-Kfz lieferten keine eindeutigen Ergebnisse. Die Analyse des Datensatzes des Expertisechecks, zeigte auffällige Abweichungen bei den Zahlen vor einer Prüfung durch einen Sachverständigen sowie bei den tatsächlich von der Versicherung ausgezahlten Beträgen. Ersteres Ergebnis belegt hierbei die Wirksamkeit des Einsatzes von Prüfgutachtern, da diese Manipulationen in den an die Versicherung übermittelten Daten aufdecken und herausstreichen. Somit ist auch von einem effizient funktionierenden Benford-Test zu erwarten, dass dieser im Falle der Beträge nach einer Prüfung eine niedrigere Wahrscheinlichkeit für vorhandene Manipulationen attestiert. Die erhöhten Abweichungen der tatsächlich ausgezahlten Beträge sind dem Abwicklungsprozess der Schadensregulierung geschuldet und liegen dem Anschein nach an der Tatsache, dass Sachbearbeiter der Versicherung im Rahmen ihrer Arbeitsrichtlinie in begründeten Fällen von dem Prüfgutachten abweichen dürfen. Sie haben demnach „das letzte Wort“ in dem Abwicklungsprozess. Im Umkehrschluss können solche Zu- oder Abschläge jedoch Verzerrungen in den

Ziffernverteilungen zur Folge haben, die mit einer erhöhten Abweichung von der Benford-Verteilung einhergehen. Eine Unterscheidung hinsichtlich des Alters des Unfall-Kfz lieferte auch bei dem Datensatz des Expertisechecks keine eindeutigen Ergebnisse.

Zusammenfassend bleibt festzuhalten, dass durch beide empirische Untersuchungen die Effektivität des Benford-Testes zum Einsatz in der forensischen Betrugserkennung sowohl in dem Themenfeld der makroökonomischen Daten als auch von Kfz-Versicherungsschäden unter Beweis gestellt wurde. Bei letzteren sollte die Anwendung bei vorhandenen Selbstbeteiligungen oder Beitragserhöhungen im Schadensfall jedoch auf die zweiten Ziffern beschränkt werden. Zwar liefert der Benford-Test selbst bei erfolgreichem Einsatz keine Beweise für vorhandene Manipulationen, jedoch kann dieser äußerst erfolgreich dazu verwendet werden, erste Hinweise für solche zu geben. Demzufolge wäre im Falle der historischen Wirtschaftsdaten der Bundesrepublik Deutschland der nächste logische Schritt, die ausgewerteten statistischen Jahrbücher insbesondere in der Zeit des Nationalsozialismus sowie der Weimarer Republik weiteren Prüfungshandlungen zu unterziehen und auf Manipulationen hin zu überprüfen. Möglicherweise würden historische Geschehnisse so neu bewertet bzw. besser nachvollzogen werden können. Ein weiterer zukünftiger Untersuchungspunkt des Benford-Testes stellen die unverdächtigen historischen Daten Deutschlands dar, welche keinen Zusammenhang mit der wirtschaftlichen Leistungsfähigkeit haben und ebenfalls in den statistischen Jahrbüchern veröffentlicht wurden. Da es bei diesen keinerlei Motivation zur Fälschung gibt, müsste der Benford-Test hier bei allen Regierungsformen eine ideale Benford-Verteilung attestieren. Ist dies der Fall, wäre dies ein weiterer Beweis zur Untermauerung der Wirksamkeit des Benford-Testes. Des Weiteren kann der Benford-Test aufgrund seiner Eignung auch empfohlen werden, zukünftig auf veröffentlichte volkswirtschaftliche Zahlen regelmäßig angewendet zu werden. Somit würden erste Anzeichen für Betrug und Falschangaben frühzeitig aufgedeckt und wirtschaftliche Fehlentwicklungen rechtzeitig erkannt. Im Themenfeld der Versicherungsdaten können die im Rahmen der Untersuchung festgestellten Ergebnisse dazu verwendet werden, diejenigen Schadensfälle mit besonders hohen Wahrscheinlichkeiten vorhandener Manipulationen zu selektieren, um diese intensiveren Prüfungen zu unterziehen. Dies kann für die Versicherungsgesellschaften enormes Potenzial für Kostensenkungen heben, da eine intensive Einzelfallprüfung überflüssig würde. Des Weiteren kann empfohlen werden, die Anwendung des Benford-Testes auf weitere Attribute auszudehnen, um mittels dieser einfachen Methodik ein umfassendes Risikoprofil für Schadensfälle zu entwickeln. Da der Benford-Test relativ einfach auch auf große Datenvolumina

anwendbar ist und dessen Applikation äußerst kostengünstig ist, bietet dieser optimale Voraussetzungen, um sich im Zeitalter steigender zur Verfügung stehender Informationsmengen als ein wesentliches Instrument zur Manipulationsaufdeckung zu etablieren. Somit birgt das Benfordsche Gesetz auch in Zukunft weitreichendes Potenzial, um die alltäglich in unserer Gesellschaft vorhandenen Manipulationsversuche wirksam und kosteneffizient zu bekämpfen.

Anhang

Anhangverzeichnis

Anhang I: Quantile der Chi-Quadratverteilung mit f Freiheitsgraden.....	165
Anhang II: Statistische Kennzahlen und Stichprobenumfänge der Einzeljahre nach Staatsform	166
Anhang III: Ergebnisse des Benford-Testes der einzelnen Staatsformen	171
Anhang IV: Einzeljahrauswertung Weimarer Republik	172
Anhang V: Einzeljahrauswertung NS-Regime	173
Anhang VI: Einzeljahrauswertung Deutsche Demokratische Republik und Bundes- republik Deutschland	174
Anhang VII: Ergebnisse des Benford-Testes bei Anwendung auf den Datensatz „Partnerwerkstätten“.....	180
Anhang VIII: Ergebnisse des Benford-Testes bei Anwendung auf den Datensatz „Expertisecheck“	186

Anhang I: Quantile der Chi-Quadrat-Verteilung mit f Freiheitsgraden

f	α								
	0,005	0,01	0,025	0,05	0,5	0,95	0,975	0,99	0,995
1	-	-	0,001	0,004	0,455	3,84	5,02	6,63	7,88
2	0,010	0,020	0,051	0,103	1,39	5,99	7,38	9,21	10,6
3	0,072	0,115	0,216	0,352	2,37	7,81	9,35	11,3	12,8
4	0,207	0,297	0,484	0,711	3,36	9,49	11,1	13,3	14,9
5	0,412	0,554	0,831	1,15	4,35	11,1	12,8	15,1	16,7
6	0,676	0,872	1,24	1,64	5,35	12,6	14,4	16,8	18,5
7	0,989	1,24	1,69	2,17	6,35	14,1	16,0	18,5	20,3
8	1,34	1,65	2,18	2,73	7,34	15,5	17,5	20,1	22,0
9	1,73	2,09	2,70	3,33	8,34	16,9	19,0	21,7	23,6
11	2,60	3,05	3,82	4,57	10,3	19,7	21,9	24,7	26,8
12	3,07	3,57	4,40	5,23	11,3	21,0	23,3	26,2	28,3
13	3,57	4,11	5,01	5,89	12,3	22,4	24,7	27,7	29,8
14	4,07	4,66	5,63	6,57	13,3	23,7	26,1	29,1	31,3
15	4,60	5,23	6,26	7,26	14,3	25,0	27,5	30,6	32,8
16	5,14	5,81	6,91	7,96	15,3	26,3	28,8	32,0	34,3
17	5,70	6,41	7,56	8,67	16,3	27,6	30,2	33,4	35,7
18	6,26	7,01	8,23	9,39	17,3	28,9	31,5	34,8	37,2
19	6,84	7,63	8,91	10,1	18,3	30,1	32,9	36,2	38,6
21	8,03	8,90	10,3	11,6	20,3	32,7	35,5	38,9	41,4
22	8,64	9,54	11,0	12,3	21,3	33,9	36,8	40,3	42,8
23	9,26	10,2	11,7	13,1	22,3	35,2	38,1	41,6	44,2
24	9,89	10,9	12,4	13,8	23,3	36,4	39,4	43,0	45,6
25	10,5	11,5	13,1	14,6	24,3	37,7	40,6	44,3	46,9
26	11,2	12,2	13,8	15,4	25,3	38,9	41,9	45,6	48,3
27	11,8	12,9	14,6	16,2	26,3	40,1	43,2	47,0	49,6
28	12,5	13,6	15,3	16,9	27,3	41,3	44,5	48,3	51,0
29	13,1	14,3	16,0	17,7	28,3	42,6	45,7	49,6	52,3

Anhang II: Statistische Kennzahlen und Stichprobenumfänge der Einzeljahre nach Staatsform

Tabelle 75: Statistische Kennzahlen und Stichprobenumfänge der Einzeljahre der Weimarer Republik:

Weimarer Republik Jahr	Stichproben- umfang n erste Ziffern	Mittelwert	Median	Schiefe
1918	4.022	1.881.446,37	989,40	21,59
1919	4.533	20.689.328,41	4.474,00	10,46
1920	4.947	3.909.106,46	1.400,00	22,35
1921/1922	5.710	6.613.218,98	3.666,55	28,25
1923	4.823	355.802.951,02	15.385,00	25,92
1924	1.951	21.408.365,52	1.382,40	41,34
1925/1926	4.144	846.568,33	3.618,00	37,40
1927	3.457	1.166.818,74	2.806,90	27,85
1928	3.671	2.954.047,21	1.843,00	35,47
1929	3.945	3.530.984,95	2.477,00	19,51
1930	4.146	3.281.533,83	1.621,85	22,33
1931	3.829	1.979.169,93	2.781,00	33,46
1932	4.011	145.229,94	1.467,00	9,92
Gesamt	53.189	30.300.626,41	3.136,58	23,99

Tabelle 76: Statistische Kennzahlen und Stichprobenumfänge der Einzeljahre des NS-Regimes:

NS-Regime	Stichproben- umfang n	Mittelwert	Median	Schiefe
Jahr	erste Ziffern			
1933	4.423	76.403,65	468,40	14,92
1934	3.109	65.273,95	929,00	11,20
1935	2.556	102.135,97	822,50	13,53
1936	2.062	53.695,55	1168,50	27,32
1937	2.015	43.054,95	717,00	10,40
1938	2.086	53.612,81	951,50	8,92
1939	2.111	72.893,47	820,00	24,62
1941/1942	2.161	66.091,89	816,80	11,45
Gesamt	20.523	59.240,25	743,74	13,60

Tabelle 77: Statistische Kennzahlen und Stichprobenumfänge der Einzeljahre der Deutschen Demokratischen Republik:

Deutsche Demo- kratische Republik	Stichproben- umfang n	Mittelwert	Median	Schiefe
Jahr	erste Ziffern			
1955	588	137.996,74	476,00	10,66
1956	937	169.244,26	28.935,00	6,77
1957	1.472	179.975,48	12.907,50	7,16
1958	1.408	337.707,85	19.792,00	16,62
1959	1.429	368.297,96	11.331,00	7,48
1960/1961	1.555	288.822,35	15.892,00	6,89
1962	1.507	331.572,86	13.085,00	6,57
1963	1.360	448.022,88	20.551,00	22,02
1964	1.485	296.773,02	11.721,00	7,06
1965	1.803	322.949,69	4.194,00	24,70
1966	2.078	271.081,50	5.055,00	28,97
1967	2.266	262.238,41	4.295,50	31,62

1968	2.059	348.466,70	6.854,00	26,44
1969	1.397	2.080.849,68	7.596,00	37,23
1970	1.369	535.513,60	7.170,20	23,01
1971	1.282	451.855,64	6.933,50	7,88
1972	1.374	583.864,28	6.959,00	7,43
1973	1.234	38.330.475,67	6.499,00	35,13
1974	1.245	410.224,31	6.940,00	9,21
1975	1.280	291.270,41	5.726,50	9,27
1976	1.303	12.273.740,33	7.781,00	36,10
1977	1.214	322.098,08	8.666,50	9,58
1978	944	352.564,52	8.597,00	9,64
1979	973	2.182.323,05	8.326,00	31,13
1980	937	379.018,18	11.366,00	9,57
1981	937	218.375,08	9.306,00	14,05
1982	998	76.703,79	5.948,00	12,63
1983	1.036	87.449,85	5.749,00	18,99
1984	1.083	77.121,07	4.397,00	12,40
1985	1.102	86.978,66	4.718,80	11,52
1986	1.141	617.769,58	4.966,40	33,76
1987	1.180	82.073,34	4.760,50	6,98
1988	1.302	76.602,56	4.218,85	6,83
1989	1.374	77.817,13	3.438,45	6,61
1990	945	354.875,98	2.082,00	30,64
Gesamt	45.597	1.820.363,27	8.492,42	16,64

Tabelle 78: Statistische Kennzahlen und Stichprobenumfänge der Einzeljahre der Bundesrepublik Deutschland:

Bundesrepublik Deutschland	Stichproben- umfang n erste Ziffern	Mittelwert	Median	Schiefe
Jahr				
1952	1.839	369.698,37	346,90	22,86
1953	1.968	346.307,60	473,90	25,08
1954	2.035	386.349,22	872,00	25,41
1955	2.362	100.885,81	859,50	12,49
1956	2.105	195.525,21	1.108,10	30,95
1957	2.549	89.223,48	928,30	27,12
1958	2.206	101.439,35	954,00	10,69
1959	1.858	146.840,27	2.045,85	8,58
1960	2.055	84.974,30	744,10	13,35
1961	2.538	112.254,81	924,25	12,87
1962	2.342	125.157,70	1.399,51	11,53
1963	2.594	144.063,61	684,00	14,39
1964	2.441	132.637,66	752,00	14,36
1965	2.568	111.449,06	615,00	14,04
1966	2.548	195.993,84	1.506,50	13,07
1967	2.942	56.436,64	910,50	15,66
1968	2.401	199.753,07	746,70	48,80
1969	2.470	39.136,15	739,25	18,99
1970	2.671	50.385,48	891,00	17,68
1971	2.574	24.191,18	1.019,95	15,58
1972	2.568	529.162,25	998,20	49,80
1973	2.796	1.254.161,21	1.142,50	37,70
1974	2.765	42.169,68	1.267,00	14,84
1975	2.547	31.729,97	1.373,00	9,50
1976	2.335	13.531,50	967,00	9,81
1977	2.677	53.332,19	1.502,80	29,91
1978	2.942	58.111,15	2.099,00	30,68

1979	2.511	61.557,18	1.806,00	28,59
1980	2.569	66.962,44	1.929,00	28,94
1981	2.525	55.648,18	2.235,00	8,57
1982	2.605	61.505,76	2.439,00	7,72
1983	2.723	68.912,63	2.648,00	9,22
1984	2.333	101.487,48	2.868,00	22,21
1985	2.737	94.575,03	2.800,20	23,14
1986	2.731	135.174,95	2.813,60	30,75
1987	2.804	94.982,94	2.947,50	26,19
1988	2.351	104.216,62	3.187,00	25,95
1989	2.735	121.521,12	3.371,00	20,60
1990	2.449	91.821,58	3.103,90	36,55
Gesamt	96.769	155.211,97	1.538,95	21,13

Anhang III: Ergebnisse des Benford-Testes der einzelnen Staatsformen

Tabelle 79: Ergebnisse des Benford-Testes der einzelnen Staatsformen bei Analyse der ersten Ziffern:

1. Ziffern	n	χ^2	KS	Kuiper	$\frac{\chi^2}{n}$	d*	a*	Manhattan	Tschebyscheff
<i>BRD</i>	96.769	25,63	0,0036	1,9698	0,0003	0,0053	0,0012	0,0141	0,0035
<i>DDR</i>	45.597	17,31	0,0050	1,4694	0,0004	0,0058	0,0014	0,0150	0,0040
<i>NS</i>	20.523	28,99	0,0089	1,4054	0,0014	0,0116	0,0053	0,0281	0,0076
<i>WR</i>	53.189	21,02	0,0062	1,5877	0,0004	0,0078	0,0025	0,0160	0,0061
<i>Gesamt</i>	199.978	4,68	0,0014	0,6484	0,0000	0,0013	0,0008	0,0033	0,0008

Tabelle 80: Ergebnisse des Benford-Testes der einzelnen Staatsformen bei Analyse der zweiten Ziffern:

2. Ziffern	n	χ^2	KS	Kuiper	$\frac{\chi^2}{n}$	d*	a*	Manhattan	Tschebyscheff
<i>BRD</i>	94.576	18,68	0,0028	1,6965	0,0002	0,0046	0,0001	0,0126	0,0027
<i>DDR</i>	44.600	10,08	0,0040	1,1987	0,0002	0,0050	0,0025	0,0128	0,0030
<i>NS</i>	19.453	7,09	0,0033	0,8118	0,0004	0,0064	0,0006	0,0174	0,0034
<i>WR</i>	51.550	61,85	0,0082	1,8667	0,0012	0,0117	0,0106	0,0297	0,0082
<i>Gesamt</i>	194.910	34,09	0,0031	1,4732	0,0002	0,0044	0,0034	0,0109	0,0029

Anhang IV: Einzeljahrauswertung Weimarer Republik

Tabelle 81: Ergebnisse des Benford-Testes bei Anwendung auf die ersten Ziffern der Weimarer Republik (Jahre mit größter Abweichung fett markiert):

Weimarer Republik 1. Ziffern	n	X^2	KS	Kuiper	$\frac{X^2}{n}$	d*	a*	Manhattan	Tschebyscheff
1918	4.022	19,82	0,0183	1,1657	0,0049	0,0238	0,0140	0,0576	0,0183
1919	4.533	30,97	0,0285	1,9259	0,0068	0,0229	0,0244	0,0574	0,0139
1920	4.947	6,16	0,0084	0,8105	0,0012	0,0092	0,0057	0,0248	0,0047
1921 und 1922	5.710	7,22	0,0060	0,5617	0,0013	0,0104	0,0046	0,0251	0,0074
1923	4.823	9,19	0,0122	1,2332	0,0019	0,0143	0,0042	0,0380	0,0098
1924	1.951	9,82	0,0284	1,2976	0,0050	0,0245	0,0244	0,0585	0,0188
1925 und 1926	4.144	9,33	0,0106	1,0877	0,0023	0,0132	0,0030	0,0337	0,0096
1927	3.457	20,77	0,0277	1,8264	0,0060	0,0275	0,0119	0,0725	0,0172
1928	3.671	11,02	0,0131	1,3158	0,0030	0,0190	0,0024	0,0482	0,0131
1929	3.945	4,43	0,0080	0,7289	0,0011	0,0107	0,0024	0,0280	0,0064
1930	4.146	12,83	0,0173	1,1199	0,0031	0,0201	0,0155	0,0435	0,0173
1931	3.829	11,26	0,0222	1,3788	0,0029	0,0204	0,0193	0,0479	0,0171
1932	4.011	8,42	0,0170	1,1513	0,0021	0,0174	0,0123	0,0395	0,0144

Tabelle 82: Ergebnisse des Benford-Testes bei Anwendung auf die zweiten Ziffern der Weimarer Republik (Jahre mit größter Abweichung fett markiert):

Weimarer Republik 2. Ziffern	n	X^2	KS	Kuiper	$\frac{X^2}{n}$	d*	a*	Manhattan	Tschebyscheff
1918	3.853	2,89	0,0065	0,6113	0,0007	0,0088	0,0006	0,0230	0,0051
1919	4.421	16,90	0,0189	1,2572	0,0038	0,0212	0,0235	0,0475	0,0165
1920	4.749	19,97	0,0130	1,3398	0,0042	0,0222	0,0018	0,0519	0,0130
1921 und 1922	5.548	13,71	0,0140	1,0416	0,0025	0,0164	0,0158	0,0415	0,0094
1923	4.735	10,00	0,0171	1,1772	0,0021	0,0149	0,0195	0,0385	0,0070
1924	1.910	20,05	0,0269	1,1807	0,0105	0,0354	0,0235	0,0812	0,0269
1925 und 1926	4.062	24,15	0,0135	1,1622	0,0059	0,0259	0,0074	0,0696	0,0135
1927	3.347	11,55	0,0146	0,9099	0,0035	0,0191	0,0140	0,0479	0,0101
1928	3.553	11,27	0,0156	0,9333	0,0032	0,0186	0,0184	0,0454	0,0112
1929	3.811	9,62	0,0181	1,1179	0,0025	0,0162	0,0156	0,0424	0,0081
1930	3.997	11,01	0,0150	1,3111	0,0028	0,0169	0,0052	0,0425	0,0092
1931	3.710	17,33	0,0148	1,0600	0,0047	0,0228	0,0149	0,0607	0,0127
1932	3.854	8,72	0,0096	0,7804	0,0023	0,0153	0,0077	0,0393	0,0100

Anhang V: Einzeljahrauswertung Nationalsozialistisches Regime

Tabelle 83: Ergebnisse des Benford-Testes bei Anwendung auf die ersten Ziffern des NS-Regimes (Jahre mit größter Abweichung fett markiert):

NS-Regime	n	χ^2	KS	Kuiper	$\frac{\chi^2}{n}$	d*	a*	Manhattan	Tschebyscheff
1. Ziffern									
1933	4.423	10,66	0,0075	0,7824	0,0024	0,0134	0,0035	0,0394	0,0065
1934	3.109	2,58	0,0026	0,2846	0,0008	0,0078	0,0012	0,0187	0,0051
1935	2.556	14,56	0,0258	1,3105	0,0057	0,0214	0,0192	0,0517	0,0145
1936	2.062	13,06	0,0173	1,3231	0,0063	0,0232	0,0053	0,0620	0,0138
1937	2.015	4,44	0,0097	0,6031	0,0022	0,0132	0,0026	0,0339	0,0079
1938	2.086	12,00	0,0124	0,8084	0,0058	0,0211	0,0013	0,0471	0,0153
1939	2.111	9,88	0,0201	0,9288	0,0047	0,0257	0,0166	0,0608	0,0201
1941 und 1942	2.161	8,21	0,0238	1,1098	0,0038	0,0187	0,0213	0,0476	0,0132

Tabelle 84: Ergebnisse des Benford-Testes bei Anwendung auf die zweiten Ziffern des NS-Regimes (Jahre mit größter Abweichung fett markiert):

NS-Regime	n	χ^2	KS	Kuiper	$\frac{\chi^2}{n}$	d*	a*	Manhattan	Tschebyscheff
2. Ziffern									
1933	4.184	12,77	0,0100	1,1852	0,0031	0,0186	0,0012	0,0423	0,0111
1934	2.953	10,17	0,0155	0,8440	0,0034	0,0193	0,0173	0,0479	0,0123
1935	2.427	10,66	0,0164	1,2179	0,0044	0,0211	0,0083	0,0505	0,0142
1936	1.962	12,34	0,0190	0,8426	0,0063	0,0270	0,0157	0,0632	0,0190
1937	1.914	18,50	0,0211	1,2675	0,0097	0,0307	0,0029	0,0763	0,0211
1938	1.985	7,65	0,0141	0,6427	0,0039	0,0207	0,0185	0,0494	0,0104
1939	1.998	16,10	0,0192	1,3699	0,0081	0,0300	0,0039	0,0759	0,0175
1941 und 1942	2.030	8,59	0,0231	1,0555	0,0042	0,0218	0,0163	0,0524	0,0152

Anhang VI: Einzeljahrauswertung Deutsche Demokratische Republik und Bundesrepublik Deutschland

Tabelle 85: Vergleich der Ergebnisse des Benford-Testes zwischen der DDR und der BRD bei Anwendung auf die ersten Ziffern (höhere Abweichung im Einzeljahr fett markiert):

BRD und DDR 1. Ziffern		<i>n</i>	χ^2	KS	Kuiper	$\frac{\chi^2}{n}$	d*	a*	Man- hattan	Tsche- by- scheff
1952	BRD	1.839	7,45	0,0247	1,0633	0,0041	0,0234	0,0124	0,0545	0,0146
1953	BRD	1.968	4,88	0,0066	0,5553	0,0025	0,0113	0,0018	0,0279	0,0069
1954	BRD	2.035	8,66	0,0149	1,0297	0,0043	0,0208	0,0061	0,0561	0,0149
1955	BRD	2.362	4,30	0,0089	0,6409	0,0018	0,0111	0,0044	0,0304	0,0067
	DDR	588	6,14	0,0314	0,7662	0,0104	0,0306	0,0309	0,0813	0,0195
1956	BRD	2.105	16,35	0,0127	0,9915	0,0078	0,0261	0,0037	0,0779	0,0128
	DDR	937	9,23	0,0225	0,9809	0,0098	0,0288	0,0076	0,0772	0,0151
1957	BRD	2.549	16,25	0,0178	1,0596	0,0064	0,0242	0,0129	0,0591	0,0148
	DDR	1.472	5,42	0,0122	0,8202	0,0037	0,0163	0,0011	0,0435	0,0104
1958	BRD	2.206	13,86	0,0241	1,3487	0,0063	0,0218	0,0153	0,0572	0,0138
	DDR	1.408	10,15	0,0192	1,0141	0,0072	0,0232	0,0079	0,0559	0,0150
1959	BRD	1.858	17,07	0,0184	1,0986	0,0092	0,0254	0,0043	0,0696	0,0131
	DDR	1.429	13,31	0,0321	1,6283	0,0093	0,0313	0,0119	0,0858	0,0221
1960	BRD	2.055	5,56	0,0226	1,0290	0,0027	0,0207	0,0162	0,0469	0,0173
	DDR*	1.555	9,93	0,0270	1,3189	0,0064	0,0280	0,0099	0,0694	0,0174
1961	BRD	2.538	12,69	0,0090	0,8955	0,0050	0,0204	0,0024	0,0511	0,0130
1962	BRD	2.342	14,22	0,0131	1,0571	0,0061	0,0213	0,0005	0,0565	0,0123
	DDR	1.507	11,75	0,0201	1,0932	0,0078	0,0242	0,0068	0,0624	0,0160
1963	BRD	2.594	5,89	0,0107	0,9589	0,0023	0,0148	0,0028	0,0399	0,0091
	DDR	1.360	11,67	0,0237	1,2547	0,0086	0,0329	0,0080	0,0844	0,0238
1964	BRD	2.441	16,01	0,0244	1,2108	0,0066	0,0278	0,0183	0,0607	0,0200
	DDR	1.485	7,43	0,0142	0,9657	0,0050	0,0213	0,0020	0,0547	0,0135

*) Ergebnisse der DDR für die Jahre 1960 und 1961 zusammengefasst.

BRD und DDR 1. Ziffern		<i>n</i>	χ^2	KS	Kuiper	$\frac{\chi^2}{n}$	d*	a*	Man- hatten	Tsche- by- scheff
1965	BRD	2.568	10,39	0,0235	1,1926	0,0040	0,0212	0,0240	0,0492	0,0172
	DDR	1.803	21,03	0,0428	1,8260	0,0117	0,0328	0,0315	0,0857	0,0169
1966	BRD	2.548	12,24	0,0207	1,3639	0,0048	0,0232	0,0057	0,0539	0,0191
	DDR	2.078	22,21	0,0325	1,4851	0,0107	0,0309	0,0207	0,0660	0,0252
1967	BRD	2.942	10,77	0,0230	1,2516	0,0037	0,0203	0,0147	0,0559	0,0128
	DDR	2.266	9,46	0,0226	1,0806	0,0042	0,0198	0,0160	0,0480	0,0124
1968	BRD	2.401	4,43	0,0149	0,7669	0,0018	0,0171	0,0073	0,0559	0,0149
	DDR	2.059	7,89	0,0171	1,0260	0,0038	0,0191	0,0166	0,0462	0,0119
1969	BRD	2.470	9,62	0,0136	0,9709	0,0039	0,0177	0,0071	0,0463	0,0104
	DDR	1.397	6,84	0,0167	0,6287	0,0049	0,0196	0,0139	0,0538	0,0111
1970	BRD	2.671	7,20	0,0098	1,0073	0,0027	0,0152	0,0003	0,0415	0,0098
	DDR	1.369	15,24	0,0101	0,6534	0,0111	0,0298	0,0004	0,0806	0,0176
1971	BRD	2.574	5,15	0,0112	0,5705	0,0020	0,0153	0,0075	0,0404	0,0112
	DDR	1.282	13,31	0,0258	1,0800	0,0104	0,0378	0,0133	0,0903	0,0258
1972	BRD	2.568	7,54	0,0148	0,7513	0,0029	0,0148	0,0125	0,0410	0,0077
	DDR	1.374	14,97	0,0097	0,6923	0,0109	0,0272	0,0010	0,0746	0,0151
1973	BRD	2.796	10,14	0,0180	0,9545	0,0036	0,0218	0,0138	0,0538	0,0180
	DDR	1.234	11,31	0,0247	1,3375	0,0092	0,0333	0,0046	0,0814	0,0247
1974	BRD	2.765	4,01	0,0060	0,5980	0,0015	0,0115	0,0024	0,0339	0,0061
	DDR	1.245	8,01	0,0161	1,1141	0,0064	0,0252	0,0019	0,0629	0,0158
1975	BRD	2.547	8,95	0,0097	0,7797	0,0035	0,0165	0,0021	0,0483	0,0088
	DDR	1.280	9,82	0,0205	1,2076	0,0077	0,0255	0,0051	0,0679	0,0130
1976	BRD	2.335	9,84	0,0197	1,4702	0,0042	0,0268	0,0014	0,0607	0,0197
	DDR	1.303	4,30	0,0164	0,7245	0,0033	0,0188	0,0052	0,0437	0,0127
1977	BRD	2.677	6,50	0,0173	0,8973	0,0024	0,0165	0,0105	0,0396	0,0111
	DDR	1.214	3,47	0,0098	0,3878	0,0029	0,0151	0,0057	0,0381	0,0089

BRD und DDR 1. Ziffern		<i>n</i>	χ^2	KS	Kuiper	$\frac{\chi^2}{n}$	<i>d</i> *	<i>a</i> *	Man- hatten	Tsche- by- scheff
1978	BRD	2.942	3,71	0,0132	0,7163	0,0013	0,0112	0,0066	0,0297	0,0064
	DDR	944	10,73	0,0239	0,9937	0,0114	0,0346	0,0051	0,0892	0,0254
1979	BRD	2.511	10,83	0,0199	1,1409	0,0043	0,0217	0,0102	0,0496	0,0175
	DDR	973	12,50	0,0330	1,3472	0,0128	0,0404	0,0160	0,1008	0,0330
1980	BRD	2.569	6,81	0,0180	0,9166	0,0027	0,0176	0,0097	0,0435	0,0115
	DDR	937	3,29	0,0234	0,7205	0,0035	0,0239	0,0231	0,0472	0,0223
1981	BRD	2.525	19,40	0,0169	1,4109	0,0077	0,0291	0,0004	0,0759	0,0207
	DDR	937	2,83	0,0129	0,6283	0,0030	0,0182	0,0074	0,0450	0,0117
1982	BRD	2.605	11,36	0,0125	1,2739	0,0044	0,0238	0,0023	0,0578	0,0155
	DDR	998	11,95	0,0206	0,9552	0,0120	0,0287	0,0060	0,0796	0,0182
1983	BRD	2.723	8,15	0,0072	0,5845	0,0030	0,0155	0,0021	0,0383	0,0104
	DDR	1.036	8,57	0,0128	0,8026	0,0083	0,0247	0,0023	0,0632	0,0150
1984	BRD	2.333	4,89	0,0104	0,6169	0,0021	0,0159	0,0046	0,0415	0,0104
	DDR	1.083	5,38	0,0157	0,6408	0,0050	0,0197	0,0128	0,0525	0,0107
1985	BRD	2.737	6,76	0,0054	0,5161	0,0025	0,0144	0,0030	0,0359	0,0091
	DDR	1.102	1,68	0,0070	0,4286	0,0015	0,0111	0,0002	0,0262	0,0070
1986	BRD	2.731	12,10	0,0085	0,7898	0,0044	0,0182	0,0038	0,0466	0,0109
	DDR	1.141	5,91	0,0140	0,7782	0,0052	0,0200	0,0012	0,0468	0,0144
1987	BRD	2.804	12,15	0,0182	1,0401	0,0043	0,0198	0,0117	0,0532	0,0124
	DDR	1.180	6,97	0,0108	0,6664	0,0059	0,0203	0,0030	0,0516	0,0127
1988	BRD	2.351	13,55	0,0284	1,3806	0,0058	0,0215	0,0206	0,0590	0,0123
	DDR	1.302	3,55	0,0122	0,5086	0,0027	0,0131	0,0067	0,0309	0,0091
1989	BRD	2.735	16,58	0,0248	1,2996	0,0061	0,0202	0,0250	0,0495	0,0114
	DDR	1.374	4,46	0,0159	0,5912	0,0032	0,0197	0,0167	0,0482	0,0156
1990	BRD	2.449	5,57	0,0070	0,5361	0,0023	0,0126	0,0016	0,0344	0,0069
	DDR	945	12,62	0,0390	1,2060	0,0134	0,0382	0,0225	0,1015	0,0227

Tabelle 86: Vergleich der Ergebnisse des Benford-Testes zwischen der BRD und der DDR bei Anwendung auf die zweiten Ziffern (höhere Abweichung im Einzeljahr fett markiert):

BRD und DDR 2. Ziffern		<i>n</i>	χ^2	KS	Kuiper	$\frac{\chi^2}{n}$	<i>d</i> *	<i>a</i> *	Man- hattan	Tsche- by- scheff
1952	BRD	1.757	15,84	0,0045	1,0850	0,0090	0,0315	0,0068	0,0835	0,0143
1953	BRD	1.873	3,28	0,0105	0,5617	0,0018	0,0133	0,0116	0,0324	0,0070
1954	BRD	1.973	4,06	0,0015	0,5444	0,0021	0,0150	0,0010	0,0387	0,0072
1955	BRD	2.285	7,86	0,0011	1,0500	0,0034	0,0187	0,0005	0,0466	0,0103
	DDR	507	3,61	0,0151	0,6339	0,0071	0,0281	0,0082	0,0724	0,0151
1956	BRD	2.053	7,68	0,0033	0,8290	0,0037	0,0198	0,0104	0,0559	0,0089
	DDR	934	13,87	0,0188	0,8386	0,0149	0,0401	0,0147	0,1060	0,0210
1957	BRD	2.503	15,49	0,0011	0,9375	0,0062	0,0266	0,0027	0,0654	0,0149
	DDR	1.449	8,31	0,0136	0,6363	0,0057	0,0246	0,0113	0,0619	0,0154
1958	BRD	2.162	9,76	0,0050	0,8214	0,0045	0,0230	0,0037	0,0541	0,0172
	DDR	1.390	7,06	0,0197	0,7365	0,0051	0,0242	0,0206	0,0574	0,0131
1959	BRD	1.822	7,40	0,0108	1,0799	0,0041	0,0210	0,0232	0,0522	0,0142
	DDR	1.409	10,53	0,0196	1,4026	0,0075	0,0277	0,0010	0,0744	0,0158
1960	BRD	1.995	4,57	0,0073	0,6146	0,0023	0,0159	0,0021	0,0397	0,0086
	DDR*	1.539	10,68	0,0157	0,6734	0,0069	0,0265	0,0130	0,0649	0,0161
1961	BRD	2.474	16,05	0,0036	1,3622	0,0065	0,0259	0,0264	0,0707	0,0127
1962	BRD	2.301	8,43	0,0021	0,8057	0,0037	0,0200	0,0146	0,0472	0,0126
	DDR	1.489	5,66	0,0147	0,9309	0,0038	0,0204	0,0003	0,0480	0,0118
1963	BRD	2.549	13,60	0,0042	1,1213	0,0053	0,0230	0,0046	0,0608	0,0107
	DDR	1.344	3,26	0,0135	0,4979	0,0024	0,0163	0,0115	0,0397	0,0095
1964	BRD	2.396	9,86	0,0036	1,0620	0,0041	0,0213	0,0070	0,0502	0,0127
	DDR	1.468	9,35	0,0138	0,7168	0,0064	0,0259	0,0101	0,0636	0,0138

*) Ergebnisse der DDR für die Jahre 1960 und 1961 zusammengefasst.

BRD und DDR 2. Ziffern		<i>n</i>	χ^2	KS	Kuiper	$\frac{\chi^2}{n}$	d*	a*	Man- hatten	Tsche- by- scheff
1965	BRD	2.518	16,95	0,0069	1,3492	0,0067	0,0276	0,0057	0,0615	0,0165
	DDR	1.768	7,91	0,0148	0,7860	0,0045	0,0213	0,0108	0,0526	0,0146
1966	BRD	2.503	7,66	0,0028	0,9825	0,0031	0,0189	0,0223	0,0425	0,0138
	DDR	2.042	5,71	0,0122	0,8517	0,0028	0,0177	0,0057	0,0446	0,0084
1967	BRD	2.894	21,35	0,0036	1,1994	0,0074	0,0290	0,0125	0,0703	0,0171
	DDR	2.233	7,89	0,0129	0,9203	0,0035	0,0189	0,0027	0,0492	0,0105
1968	BRD	2.361	11,90	0,0027	0,9688	0,0050	0,0230	0,0002	0,0610	0,0132
	DDR	2.037	7,50	0,0087	0,6851	0,0037	0,0198	0,0056	0,0495	0,0120
1969	BRD	2.430	16,08	0,0022	1,3357	0,0066	0,0268	0,0021	0,0672	0,0175
	DDR	1.379	9,46	0,0189	0,9278	0,0069	0,0280	0,0058	0,0671	0,0203
1970	BRD	2.623	11,56	0,0063	0,9221	0,0044	0,0210	0,0015	0,0407	0,0179
	DDR	1.356	11,32	0,0297	1,1987	0,0083	0,0300	0,0290	0,0733	0,0195
1971	BRD	2.530	13,32	0,0015	1,1253	0,0053	0,0231	0,0020	0,0604	0,0121
	DDR	1.274	5,28	0,0114	0,4100	0,0041	0,0204	0,0141	0,0553	0,0100
1972	BRD	2.501	6,38	0,0032	0,7473	0,0026	0,0165	0,0111	0,0360	0,0101
	DDR	1.360	8,29	0,0137	0,7684	0,0061	0,0250	0,0061	0,0693	0,0139
1973	BRD	2.730	14,10	0,0078	1,2767	0,0052	0,0249	0,0134	0,0487	0,0188
	DDR	1.222	15,12	0,0215	1,3150	0,0124	0,0357	0,0022	0,0827	0,0228
1974	BRD	2.699	14,08	0,0025	1,2075	0,0052	0,0237	0,0051	0,0592	0,0150
	DDR	1.236	7,39	0,0090	0,5696	0,0060	0,0256	0,0041	0,0661	0,0158
1975	BRD	2.490	11,69	0,0050	1,2013	0,0047	0,0233	0,0127	0,0546	0,0145
	DDR	1.259	15,21	0,0181	0,9817	0,0121	0,0357	0,0119	0,0807	0,0222
1976	BRD	2.251	12,53	0,0064	0,9717	0,0056	0,0252	0,0148	0,0627	0,0178
	DDR	1.285	17,62	0,0171	1,1155	0,0137	0,0373	0,0113	0,0958	0,0169
1977	BRD	2.592	5,01	0,0007	0,8677	0,0019	0,0145	0,0059	0,0342	0,0076
	DDR	1.196	9,11	0,0337	1,1805	0,0076	0,0290	0,0302	0,0719	0,0160

BRD und DDR 2. Ziffern		<i>n</i>	χ^2	KS	Kuiper	$\frac{\chi^2}{n}$	d*	a*	Man- hattan	Tsche- by- scheff
1978	BRD	2.863	9,18	0,0178	0,9893	0,0032	0,0187	0,0151	0,0508	0,0085
	DDR	882	12,80	0,0286	0,8549	0,0145	0,0394	0,0265	0,0992	0,0246
1979	BRD	2.455	9,21	0,0118	0,7973	0,0038	0,0200	0,0122	0,0416	0,0138
	DDR	915	5,20	0,0230	1,0143	0,0057	0,0248	0,0090	0,0667	0,0126
1980	BRD	2.509	9,69	0,0094	0,8093	0,0039	0,0202	0,0080	0,0496	0,0118
	DDR	889	11,91	0,0255	1,2608	0,0134	0,0388	0,0120	0,1012	0,0229
1981	BRD	2.476	9,92	0,0111	0,6091	0,0040	0,0201	0,0084	0,0545	0,0111
	DDR	898	13,18	0,0287	0,8672	0,0147	0,0400	0,0196	0,0993	0,0204
1982	BRD	2.547	5,97	0,0111	0,6693	0,0023	0,0159	0,0036	0,0361	0,0100
	DDR	962	5,82	0,0195	0,9720	0,0061	0,0247	0,0059	0,0668	0,0133
1983	BRD	2.662	17,57	0,0182	1,5740	0,0066	0,0273	0,0089	0,0656	0,0173
	DDR	1.005	11,55	0,0351	1,3162	0,0115	0,0346	0,0288	0,0826	0,0217
1984	BRD	2.286	9,25	0,0225	1,0809	0,0040	0,0202	0,0231	0,0538	0,0099
	DDR	1.048	4,94	0,0093	0,5021	0,0047	0,0221	0,0070	0,0635	0,0104
1985	BRD	2.679	11,46	0,0199	1,3087	0,0043	0,0212	0,0189	0,0563	0,0110
	DDR	1.070	8,48	0,0187	0,6147	0,0079	0,0293	0,0178	0,0742	0,0144
1986	BRD	2.668	7,79	0,0107	0,9327	0,0029	0,0177	0,0030	0,0419	0,0103
	DDR	1.097	8,43	0,0142	0,8036	0,0077	0,0281	0,0056	0,0622	0,0195
1987	BRD	2.747	3,22	0,0068	0,5605	0,0012	0,0111	0,0030	0,0288	0,0056
	DDR	1.143	5,10	0,0117	0,7179	0,0045	0,0221	0,0029	0,0623	0,0111
1988	BRD	2.319	5,99	0,0133	0,8711	0,0026	0,0161	0,0074	0,0433	0,0080
	DDR	1.265	5,27	0,0114	0,4878	0,0042	0,0216	0,0100	0,0589	0,0114
1989	BRD	2.693	15,73	0,0212	1,5286	0,0058	0,0253	0,0112	0,0637	0,0145
	DDR	1.338	8,71	0,0205	0,7576	0,0065	0,0269	0,0250	0,0693	0,0178
1990	BRD	2.407	14,84	0,0165	1,1349	0,0062	0,0258	0,0104	0,0722	0,0138
	DDR	912	4,41	0,0229	0,6955	0,0048	0,0220	0,0230	0,0549	0,0115

Anhang VII: Ergebnisse des Benford-Testes bei Anwendung auf den Datensatz „Partnerwerkstätten“

Tabelle 87: Ergebnisse des Benford-Testes der zweiten Ziffern bei Unterscheidung nach Art der Werkstatt:

Ergebnisse 2. Ziffern	n	χ^2	KS	Kui- per	$\frac{\chi^2}{n}$	d*	a*	Man- hattan	Tscheby- scheff
2013									
<i>Markenwerk- stätten</i>	2.550	6,10	0,0112	0,7466	0,0024	0,0156	0,0087	0,0364	0,0089
<i>Freie Werkstätten</i>	1.278	12,85	0,0302	1,0862	0,0101	0,0329	0,0230	0,0752	0,0209
2014									
<i>Markenwerk- stätten</i>	3.463	10,22	0,0179	1,0563	0,0030	0,0183	0,0224	0,0456	0,0125
<i>Freie Werkstätten</i>	1.645	14,16	0,0336	1,3672	0,0086	0,0306	0,0374	0,0813	0,0152

Tabelle 88: Ergebnisse der Analyse der zweiten Ziffern bei Unterscheidung nach Kfz-Hersteller

Ergebnisse 2. Ziffern	n	χ^2	KS	Kui- per	$\frac{\chi^2}{n}$	d*	a*	Man- hattan	Tscheby- scheff
2013									
<i>Premiumherstel- ler A</i>	350	3,13	0,0114	0,3899	0,0089	0,0312	0,0050	0,0787	0,0174
<i>Premiumherstel- ler B</i>	383	14,20	0,0458	1,0355	0,0371	0,0619	0,0229	0,1275	0,0502
<i>Premiumherstel- ler C</i>	442	12,20	0,0531	1,4168	0,0276	0,0549	0,0480	0,1392	0,0279
<i>Massenhersteller A</i>	267	12,69	0,0564	1,2796	0,0475	0,0713	0,0093	0,1855	0,0419
<i>Massenhersteller B</i>	406	8,01	0,0236	0,8117	0,0197	0,0470	0,0000	0,1203	0,0266
<i>Massenhersteller C</i>	117	5,46	0,0337	0,6524	0,0466	0,0712	0,0194	0,1840	0,0391
<i>Massenhersteller D</i>	838	12,72	0,0353	1,3515	0,0152	0,0403	0,0357	0,1001	0,0233
2014									
<i>Premiumherstel- ler A</i>	491	7,23	0,0185	0,8089	0,0147	0,0391	0,0034	0,1002	0,0268
<i>Premiumherstel- ler B</i>	520	11,80	0,0413	1,3540	0,0227	0,0488	0,0302	0,1209	0,0297
<i>Premiumherstel- ler C</i>	542	14,78	0,0601	1,4092	0,0273	0,0559	0,0584	0,1367	0,0401
<i>Massenhersteller A</i>	363	12,63	0,0586	1,1272	0,0348	0,0607	0,0773	0,1558	0,0335
<i>Massenhersteller B</i>	519	14,94	0,0375	0,8594	0,0288	0,0554	0,0399	0,1436	0,0287
<i>Massenhersteller C</i>	194	4,93	0,0448	0,4922	0,0254	0,0499	0,0103	0,1129	0,0319
<i>Massenhersteller D</i>	1.096	9,61	0,0284	0,5464	0,0088	0,0306	0,0035	0,1267	0,0285

Tabelle 89: Ergebnisse der Analyse der zweiten Ziffern der freien Werkstätten bei Unterscheidung nach Kfz-Hersteller:

Ergebnisse 2. Ziffern	n	χ^2	KS	Kui- per	$\frac{\chi^2}{n}$	d*	a*	Man- hattan	Tscheby- scheff
2013									
<i>Premiumherstel- ler A</i>	117	9,63	0,0399	0,7504	0,0823	0,0972	0,0108	0,2346	0,0575
<i>Premiumherstel- ler B</i>	162	14,03	0,0696	1,1986	0,0866	0,0940	0,0290	0,2095	0,0733
<i>Premiumherstel- ler C</i>	108	6,79	0,0554	0,6903	0,0629	0,0814	0,0592	0,2056	0,0491
<i>Massenhersteller B</i>	125	13,39	0,0674	1,2384	0,1071	0,1067	0,0375	0,2707	0,0592
<i>Massenhersteller D</i>	262	11,03	0,0319	0,7222	0,0421	0,0667	0,0293	0,1651	0,0356
2014									
<i>Premiumherstel- ler A</i>	141	7,28	0,0492	0,7621	0,0516	0,0761	0,0318	0,2038	0,0401
<i>Premiumherstel- ler B</i>	170	16,15	0,0823	1,5024	0,0950	0,0988	0,0650	0,2541	0,0685
<i>Premiumherstel- ler C</i>	117	10,69	0,0518	0,6631	0,0914	0,0986	0,0428	0,2558	0,0572
<i>Massenhersteller B</i>	138	12,09	0,0454	0,8450	0,0876	0,0950	0,0288	0,2354	0,0519
<i>Massenhersteller D</i>	360	7,16	0,0313	1,1446	0,0199	0,0464	0,0038	0,1196	0,0283

Tabelle 90: Ergebnisse der Analyse der zweiten Ziffern der Markenwerkstätten bei Unterscheidung nach Kfz-Hersteller:

Ergebnisse 2. Ziffern	n	χ^2	KS	Kui- per	$\frac{\chi^2}{n}$	d*	a*	Man- hattan	Tscheby- scheff
2013									
<i>Premiumherstel- ler A</i>	233	7,77	0,0238	0,6451	0,0334	0,0594	0,0021	0,1591	0,0370
<i>Premiumherstel- ler B</i>	221	6,42	0,0283	0,5994	0,0290	0,0569	0,0184	0,1274	0,0370
<i>Premiumherstel- ler C</i>	334	15,51	0,0575	1,3381	0,0464	0,0696	0,0444	0,1736	0,0442
<i>Massenhersteller B</i>	281	6,67	0,0253	0,8285	0,0237	0,0505	0,0168	0,1277	0,0245
<i>Massenhersteller D</i>	576	17,00	0,0546	1,3583	0,0295	0,0558	0,0653	0,1398	0,0337
2014									
<i>Premiumherstel- ler A</i>	350	6,96	0,0240	0,7926	0,0199	0,0452	0,0081	0,1160	0,0266
<i>Premiumherstel- ler B</i>	350	9,45	0,0228	0,7697	0,0270	0,0526	0,0133	0,1332	0,0348
<i>Premiumherstel- ler C</i>	425	15,04	0,0689	1,5456	0,0354	0,0638	0,0628	0,1576	0,0433
<i>Massenhersteller B</i>	381	9,96	0,0453	0,8914	0,0261	0,0525	0,0440	0,1407	0,0252
<i>Massenhersteller D</i>	736	4,27	0,0163	0,7503	0,0058	0,0247	0,0034	0,0619	0,0161

Tabelle 91: Ergebnisse des Benford-Testes bei Anwendung auf die zweiten Ziffern bei Unterscheidung nach Art des Schadens:

Ergebnisse 2. Ziffern	n	χ^2	KS	Kui- per	$\frac{\chi^2}{n}$	d*	a*	Man- hattan	Tscheby- scheff
2013									
<i>Kasko-Schaden</i>	2.712	14,74	0,0161	1,2817	0,0054	0,0247	0,0116	0,0631	0,0157
<i>Haftpflicht- Schaden</i>	1.116	10,05	0,0202	0,8401	0,0090	0,0313	0,0216	0,0683	0,0183
2014									
<i>Kasko-Schaden</i>	3.523	22,98	0,0344	2,0482	0,0065	0,0271	0,0438	0,0688	0,0160
<i>Haftpflicht- Schaden</i>	1.585	10,35	0,0206	1,3220	0,0065	0,0270	0,0097	0,0688	0,0137

Tabelle 92: Ergebnisse des Benford-Testes bei Anwendung auf die zweiten Ziffern der freien Werkstätten bei Unterscheidung nach Art des Schadens:

Ergebnisse 2. Ziffern	n	χ^2	KS	Kui- per	$\frac{\chi^2}{n}$	d*	a*	Man- hattan	Tscheby- scheff
2013									
<i>Kasko-Schaden</i>	949	14,29	0,0360	1,1138	0,0151	0,0404	0,0377	0,0972	0,0225
<i>Haftpflicht-Schaden</i>	329	10,50	0,0406	1,1202	0,0319	0,0580	0,0194	0,1586	0,0298
2014									
<i>Kasko-Schaden</i>	1.166	14,33	0,0439	1,5073	0,0123	0,0364	0,0550	0,0979	0,0171
<i>Haftpflicht-Schaden</i>	479	12,30	0,0327	1,0885	0,0257	0,0515	0,0055	0,0987	0,0453

Tabelle 93: Ergebnisse des Benford-Testes bei Anwendung auf die zweiten Ziffern der Markenwerkstätten bei Unterscheidung nach Art des Schadens:

Ergebnisse 2. Ziffern	n	χ^2	KS	Kui- per	$\frac{\chi^2}{n}$	d*	a*	Man- hattan	Tscheby- scheff
2013									
<i>Kasko-Schaden</i>	1.763	16,32	0,0093	1,1907	0,0093	0,0314	0,0025	0,0807	0,0152
<i>Haftpflicht-Schaden</i>	787	12,10	0,0154	0,8848	0,0154	0,0412	0,0226	0,1003	0,0243
2014									
<i>Kasko-Schaden</i>	2.357	15,18	0,0318	1,5483	0,0064	0,0269	0,0383	0,0686	0,0167
<i>Haftpflicht-Schaden</i>	1.106	7,95	0,0157	1,0402	0,0072	0,0293	0,0115	0,0640	0,0199

Tabelle 94: Ergebnisse des Benford-Testes bei Anwendung auf die zweiten Ziffern bei Unterscheidung nach Datum der Erstzulassung

Ergebnisse 2. Ziffern	n	χ^2	KS	Kui- per	$\frac{\chi^2}{n}$	d*	a*	Man- hattan	Tscheby- scheff
2013									
< 2000	228	10,95	0,0524	1,0180	0,0480	0,0703	0,0207	0,1600	0,0444
2000 – 2002	238	4,18	0,0244	0,4126	0,0176	0,0435	0,0189	0,1123	0,0257
2003 – 2005	526	8,45	0,0229	0,5291	0,0161	0,0421	0,0227	0,1048	0,0229
2006 – 2008	951	8,73	0,0222	0,9114	0,0092	0,0319	0,0194	0,0886	0,0177
2009 - 2011	1.290	15,48	0,0176	0,9331	0,0120	0,0363	0,0058	0,0945	0,0187
> 2012	502	10,06	0,0264	0,9043	0,0200	0,0469	0,0203	0,1007	0,0348
2014									
< 2001	357	3,67	0,0256	0,6890	0,0103	0,0335	0,0129	0,0876	0,0178
2001 – 2003	361	14,37	0,0682	1,4488	0,0398	0,0666	0,0578	0,1856	0,0314
2004 – 2006	762	11,40	0,0406	1,1261	0,0150	0,0413	0,0410	0,0999	0,0252
2007 – 2009	1.295	12,08	0,0274	0,9910	0,0093	0,0329	0,0365	0,0690	0,0247
2010 - 2012	1.558	17,10	0,0266	1,2738	0,0110	0,0342	0,0233	0,0894	0,0208
> 2013	629	5,42	0,0185	0,7764	0,0086	0,0308	0,0113	0,0717	0,0185

Anhang VIII: Ergebnisse des Benford-Testes bei Anwendung auf den Datensatz „Expertisecheck“

Tabelle 95: Ergebnisse der zweiten Ziffern des Datensatzes „Expertisecheck“ bei Unterscheidung nach Stand der Prüfung

Ergebnisse 2. Ziffern	n	χ^2	KS	Kui- per	$\frac{\chi^2}{n}$	d*	a*	Man- hattan	Tscheby- scheff
2013									
<i>Vor der Prüfung</i>	17.128	31,61	0,0137	1,7937	0,0018	0,0149	0,0185	0,0300	0,0127
<i>Nach der Prüfung</i>	17.127	9,20	0,0060	1,1468	0,0005	0,0077	0,0031	0,0214	0,0035
<i>Tatsächliche Ausschüttung</i>	17.125	31,87	0,0187	2,4489	0,0019	0,0147	0,0206	0,0374	0,0100
2014									
<i>Vor der Prüfung</i>	21.085	39,07	0,0187	2,7239	0,0019	0,0144	0,0218	0,0375	0,0088
<i>Nach der Prüfung</i>	21.083	25,54	0,0075	1,8405	0,0012	0,0118	0,0005	0,0277	0,0075
<i>Tatsächliche Ausschüttung</i>	21.077	43,49	0,0181	2,6242	0,0021	0,0155	0,0220	0,0361	0,0113

Tabelle 96: Ergebnisse des Benford-Testes bei Anwendung auf die zweiten Ziffern des Datensatzes „Expertisecheck“ bei Unterscheidung nach Datum der Erstzulassung

Ergebnisse 2. Ziffern	n	χ^2	KS	Kui- per	$\frac{\chi^2}{n}$	d*	a*	Man- hattan	Tscheby- scheff
2013									
< 2000	1.948	9,38	0,0108	0,6627	0,0048	0,0227	0,0039	0,0605	0,0143
2000 – 2002	1.494	22,60	0,0396	1,5376	0,0151	0,0391	0,0488	0,0932	0,0241
2003 – 2005	2.368	20,32	0,0246	1,2019	0,0086	0,0314	0,0276	0,0832	0,0162
2006 – 2008	3.587	10,56	0,0165	0,9926	0,0029	0,0190	0,0160	0,0391	0,0165
2009 - 2011	4.776	17,77	0,0267	1,8497	0,0037	0,0204	0,0302	0,0536	0,0104
> 2012	2.094	11,08	0,0141	1,0019	0,0053	0,0247	0,0029	0,0601	0,0141
2014									
< 2001	2.366	4,91	0,0068	0,5601	0,0021	0,0150	0,0015	0,0424	0,0068
2001 – 2003	1.781	11,00	0,0296	1,2747	0,0062	0,0251	0,0297	0,0656	0,0128
2004 – 2006	3.176	5,48	0,0092	0,5178	0,0017	0,0142	0,0093	0,0334	0,0092
2007 – 2009	4.593	29,53	0,0304	2,0632	0,0064	0,0268	0,0369	0,0646	0,0163
2010 - 2012	5.676	20,31	0,0246	1,8566	0,0036	0,0199	0,0222	0,0551	0,0094
> 2013	2.457	10,98	0,0255	1,2668	0,0045	0,0220	0,0317	0,0595	0,0110

- Alt, Raimund**, Statistik – Eine Einführung für Wirtschaftswissenschaftler, 1. Aufl., Wien 2010.
- Baetge, Jörg/ Melcher, Thorsten/ Schulz, Roland**, Vermeidung von Bilanzdelikten durch (Früh-) Erkennungsmethoden-Trends in der Wirtschaftsprüfung, in: George Herde (Hrsg.), 3. Deggendorfer Forum zur Digitalen Datenanalyse, Bilanzdelikte erkennen und vermeiden, Deggendorf 2008, S. 25-54.
- Bankhofer, Udo/ Vogel, Jürgen**, Datenanalyse und Statistik, Wiesbaden 2008.
- Bauer, Johannes/ Gross, Jochen**, Difficulties in Detecting Fraud? The Use of Benford's Law on Regression Tables, in: Journal of Economics and Statistics, 2011, S. 733-748.
- Beber, B./ Scacco, A.**, What the Numbers Say: A Digit-Based Test for Election Fraud, in: Political Analysis, 20. Jg., Nr. 2, S. 211-234.
- Benford, Frank**, The Law of Anomalous Numbers, in: Proceedings of the American Philosophical Society, 1938, S. 551-572.
- Blenkers, Michael**, Chi-Test – oder „Jeder Mensch hat seine Lieblingszahl“, in: Die steuerliche Betriebsprüfung, 2003, S. 261-264.
- Bleymüller, Josef/ Gehlter, Günther/ Güblicher, Herbert**, Statistik für Wirtschaftswissenschaftler, 15. Aufl., München 2008.
- Bomdsdorf, Eckart**, Wahrscheinlichkeitsrechnung und statistische Inferenz, 8. Aufl., Köln 2002.
- Boyle, Jeff**, An Application of Fourier Series to the most Significant Digit Problem, in: The American Mathematical Monthly, 1994, S. 879-886.
- Breunig, C./ Goerres, A.**, Searching for electoral irregularities in an established democracy: Applying Benford's Law tests to Bundestag elections in Unified Germany, in: Electoral Studies, 2011, 30. Jg., Nr. 3, S. 534-545.
- Busta, Bruce/ Weinberg, Randy**, Using Benford's law and neural networks as review procedure, in: Managerial Auditing Journal, 1998, S. 356-366.
- Bücker, Rüdiger**, Statistik für Wirtschaftswissenschaftler, 5. Aufl., Oldenburg, 2003.
- Büning, Herbert/ Trenkler, Götz**, Nichtparametrische statistische Methoden, 2. Aufl., Berlin et al. 1998.

- Carslaw, Charles A. P. N.**, Anomalies in Income Numbers: Evidence of Goal Oriented Behaviour, in: The Accounting Review, 1988, S. 321-327.
- Cho, Wendy K. Tam/ Gaines, Brian J.**, Breaking the (Benford) Law: Statistical Fraud Detection in Campaign Finance, in: The American Statistician, 2007, S. 221-223.
- Cleary, Richard/ Thibodeau, Jay C.**, Applying Digital Analysis Using Benford's Law to Detect Fraud: The Dangers of Type I Errors, in: Auditing: A Journal Of Practice & Theory, Vol. 24, No. 1, 2005, S. 79.
- De Ceuster, Marc J. K./ Dhaene, Geert/ Schateman, Tom**, On the hypothesis of psychological barriers in stock markets and Benford's Law, in: Journal of Empirical Finance, 1998, S. 263-279.
- Deckert, J./ Myagkov, M./ Ordeshook, P.**, Benford's Law and the Detection of Election Fraud, in: Political Analysis, 2011, S. 245-268.
- Degen, Horst/ Lorscheid, Peter**, Statistik-Lehrbuch mit Wirtschafts- und Bevölkerungsstatistik, 2. Aufl., München/Wien 2002.
- Diaconis, Persi/ Freedman, David**, On Rounding Percentages, in: Journal of the American Statistical Association, Vol. 74, No. 366, 1979, S. 359-364.
- Diekmann, Andreas**, Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data, in: Journal of Applied Statistics, 2007, S. 321-329.
- Diekmann, Andreas/ Jann, Ben**, Benford's Law and Fraud Detection. Facts and Legends, in: German Economic Review, Vol. 11, No. 3, 2010, S. 397 – 401.
- Durtschi, Cindy/ Hillison, William/ Pacini, Carl**, The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data, Journal of Forensic Accounting, 2004, Jg. 5, Nr. 1, S. 17-34.
- Eichholtz, Dietrich**, Geschichte der Deutschen Kriegswirtschaft 1939 – 1945, München 2003.
- Eckstein, Peter, P.**, Repetitorium Statistik, 6. Aufl., Wiesbaden 2006.
- Frey, Herbert**, Monte Carlo-Simulation – Quantitative Risikoanalyse für die Versicherungsindustrie, München 2011.
- Giles, David E.**, Benford's Law and naturally occurring prices in certain eBay auctions, in: Applied Economic Letters, 2007, S. 157-161.

- Gonzalez-Garcia, Jesus/ Pastor, Gonzalo**, Benford's Law and Macroeconomic Data Quality, in: International Monetary Fund, Working Paper, 2009.
- Götze, Wolfgang/ Deutschmann, Christel/ Link, Heike**, Statistik, München 2002, S. 331.
- Guan, Liming/ He, Daoping/ Yang, David**, Auditing, integral approach to quarterly reporting, and cosmetic earnings management, in: Managerial Auditing Journal, 2006, S. 569-581.
- Günnel, Stefan/ Tödter, Karl-Heinz**, Does Benford's Law hold in economic research and forecasting?, in: Empirica, 2009, S. 273-292.
- Hafner, Robert**, Nichtparametrische Verfahren der Statistik, Wien 2001, S. 182.
- Hartmann, Stephan**, Simulation, in: Mittelstraß, Jürgen [Hrsg.], Enzyklopädie Philosophie und Wissenschaftstheorie, Stuttgart 1995, Bd. 3.
- Healy, Paul M./ Wahlen, James M.**, A Review of the Earnings Management Literature and Its Applications for Standard Setting, in: Accounting Horizons, 1999, S. 365-383.
- Hedderich, Jürgen/Sachs, Lothar**, Angewandte Statistik, 14. Aufl., Berlin/Heidelberg 2012.
- Henselmann, Klaus/ Scherr, Elisabeth/ Ditter, Dominik**, Applying Benford's Law to individual financial reports: An empirical investigation on the basis of SEC XBRL filings, Nürnberg 2012.
- Hill, Theodore P.**, Random-Number guessing and the first digit phenomenon, in: Psychological Reports, 1988, S. 967-971.
- Hill, Theodore P.**, A statistical derivation of the significant digit law, in: Statistical Science, Vol. 10, No. 4, 1995, S. 354-363.
- Hill, Theodore P.**, Base-Invariance Implies Benford's Law, in: Proceedings of the American Mathematical Society, Vol. 123, No. 3, 1995, S. 887-895.
- Hochstädter, Dieter**, Statistische Methodenlehre, 8. Aufl., Frankfurt a.M. 1996.
- Judge, George/ Schechter, Laura**, Detecting Problems in Survey Data Using Benford's Law, in: The Journal of Human Resources, 2009, S. 1-24.
- Kienle, Stephan**, What Benford Can Tell Us About Cover Pools – An Empirical Analysis, in: International Business & Economics Research Journal, Vol. 14, No. 6, 2015, S. 829-833.

- Kinnunen, Juha/ Koskela, Markku**, Who Is Miss World in Cosmetic Earnings Management? A Cross-National Comparison of Small Upward Rounding of Net Income Numbers among Eighteen Countries, in: Journal of International Accounting Research, 2003, S. 39-68.
- Kolonko, Michael**, Stochastische Simulation – Grundlagen, Algorithmen und Anwendungen, Wiesbaden 2008.
- Koxholt, Rolf**, Die Simulation - ein Hilfsmittel der Unternehmensforschung, München/Wien 1967.
- Krämer, Walter**, The Cult of Statistical Significance – What Economists Should and Should Not Do to Make their Data Talk, in: Schmollers Jahrbuch, 2011 S. 455 - 468.
- Kuntluru, Sudershan/ Shette, Rachappa/ Achalapathi, K. V.**, Applying Benford's Law to Examine the Quality of Reported Income Numbers of Unlisted Firms, in: International Journal of Economics and Finance, Vol. 7, No. 12, 2015, S. 211-218.
- Leemis, Lawrence M./ Schmeiser, Bruce W./ Evans, Diane L.**, Survival Distributions Satisfying Benford's Law, in: American Statistician, 2000, S. 236-241.
- Ley, Eduardo/ Varian, Hal R.**, Are there psychological barriers in the Dow-Jones index?, in: Applied Financial Economics, 1994, S. 217-224.
- Ley, Eduardo**, On the Peculiar Distribution of the U.S. Stock Indexes' Data, in: The American Statistician, 1996, S. 311-313.
- Mebane, W.**, Election forensics: The second-digit Benford's law test and recent American presidential elections, in: Election Fraud: Detecting and Deterring Electoral Manipulation, edited by R. Alvarez, T. Hall und S. Hyde, Washington DC, 2008, S. 161-181.
- Mebane, W.**, Fraud in the 2009 presential election in Iran?, in: Chance, 2010, S. 6-15.
- Mebane, W.**, Comment on "Benford's Law and the Detection of Election Fraud", in: Political Analysis, 2011, S. 269-272.
- Michalski T./ Stoltz G.**, Do countries falsify economic data strategically? Some evidence that they might, in: Review of Economics and Statistics, Vol. 94, S. 1-83.
- Mochty, Ludwig**, Die Aufdeckung von Manipulationen im Rechnungswesen – Was leistet Benford's Law?, in: Die Wirtschaftsprüfung, 2002, 725-736.

- Mollat, Ingolf**, Benford reloaded – IT-Forensik: EY und Fraunhofer SIT gelingt der Durchbruch, in: ZRFC, 2014, S. 14-17.
- Müller-Gronbach, Thomas/Novak, Erik/Ritter, Klaus**, Monte Carlo-Algorithmen, Berlin Heidelberg 2012.
- Nagasaka, Kenji**, On Benford's Law, in: Annals of the Institute of Statistical Mathematics, 1984, S. 337-352.
- Newcomb, Simon**, Note on the Frequency of Use of the Different Digits in Natural Numbers, in: American Journal of Mathematics, 1881, S. 39-40.
- Nigrini, Mark J.**, A Taxpayer Compliance Application of Benford's Law, in: The Journal of the American Taxation Association, 1996, S. 72-91.
- Nigrini Mark J.**, Benford's Law, New Jersey 2012.
- Nigrini, Mark J.**, Digital Analysis Using Benford's Law: Test & Statistics for Auditors, 2. Aufl., Vancouver 2000.
- Nigrini, Mark J.**, The detection of income evasion through an analysis of digital distributions, Dissertation, Department of Accounting, Univ. Cincinnati, Cincinnati OH.
- Nigrini, Mark/ Miller, Steven**, Benford's Law applied to Hydrological Data-Results and Relevance to other Geophysical Data, in: Journal of Mathematical Geology, 2007, Nr. 5, S. 469-490.
- Nigrini, Mark J./ Mittermaier, Linda J.**, The Use of Benford's Law as an Aid in Analytical Procedures, 1997, S. 52-67.
- Niskanen, Jyrki/ Keloharju, Matti**, Earnings cosmetics in a tax-driven accounting environment: evidence from Finnish public firms, in: The European Accounting Review, 2000, S. 443-452.
- Nye, John/ Moul, Charles**, The Political Economy of Numbers: On the Application of Benford's Law to International Macroeconomic Statistics, in: The B.E. Journal of Macroeconomics, 2007, S. 1-12.
- Odenthal, Roger**, Digitale Ziffernanalyse: Ein wirkungsvoller Beitrag zur computergetstützten Deliktrevision?, in: Die Wirtschaftsprüfung, 1999, S. 630-635.
- Odenthal, Roger**, Defraudanten auf der Spur, in: Betrieb und Wirtschaft, 2005, S. 1019-1023.

- Pinkham, Roger S.**, On the Distribution of First Significant Digits, in: The Annals of Mathematical Statistics, Vol. 32, No. 4, 1961, S. 1223-1230.
- Polasek, Wolfgang**, Schließende Statistik, Berlin, 1997.
- Posch, Peter, N.**, Ziffernanalyse in Theorie und Praxis, Testverfahren zur Fälschungsspürung mit Benfords Gesetz, Aachen, 2005.
- Quick, Reiner/ Wolz. Matthias**, Benford's Law in deutschen Rechnungslegungsdaten, in: Betriebswirtschaftliche Forschung und Praxis, 2003, S. 208-224.
- Rauch, Bernhard/ Götttsche, Max/ Brähler, Gernot/ Brinkert, Daniel/ Kronfeld, Thomas**, European Integration and Statistical Credibility: Application of Benford's Law on Governmental Economic Data of EU Membership Aspirants, in: The Empirical Economics Letters, 2014, Vol. 13, No. 7, S. 755 – 762.
- Rauch, Bernhard/ Götttsche, Max/ Brähler, Gernot/ Engel, Stefan**, Fact and Fiction in EU-Governmental Economic Data, in: German Economic Review, 2011, S. 243-255.
- Rauch, Bernhard/ Götttsche, Max/ Brähler, Gernot/ Geidel, Felix/ Pietras, Thomas**, Überprüfung der Rechenschaftsberichte deutscher Parteien mit Hilfe des Benford's Law, in: BFuP, 2014, S. 175-191.
- Rauch, Bernhard/ Götttsche, Max/ Brähler, Gernot/ Kronfeld, Thomas**, Deficit versus social statistics: empirical evidence for the effectiveness of Benford's law, in: Applied Economic Letters, 2014, Vol. 21, No. 3, S. 147-151.
- Rauch, Bernhard/ Götttsche, Max/ Langenegger, Stephan**, Detecting Problems in Military Expenditure Data Using Digital Analysis, in: Defence and Peace Economics, 2014, Vol. 25, No. 2, S. 97-111.
- Rodriguez, Ricardo J.**, First Significant Digit Patterns from Mixtures of Uniform Distributions, in: The American Statistician, 2004, S. 64-71.
- Rodriguez, Ricardo J.**, Reducing False Alarms in the Detection of Human Influence on Data, in Journal of Accounting, Auditing & Finance, 2004, S. 141-158.
- Ross, Sheldon M.**, Simulation, 5. Aufl., San Diego 2013.
- Rößler, Irene/ Ungerer, Albrecht**, Statistik für Wirtschaftswissenschaftler, 2. Aufl., Heidelberg 2011.
- Schatte, Peter**, On mantissa distributions in computing and Benford's Law, in: Journal of information processing and cybernetics, 1988, S. 443-455.

- Schäfer, Christin/ Schräpler, Jörg-Peter/ Müller, Klaus-Robert/ Wagner, Gert G.,** Automatic Identification of Faked and Fraudulent Interviews in the German SOEP, in: Schmollers Jahrbuch, 2005, S. 183-193.
- Schlittgen, Rainer,** Multivariate Statistik, München 2009.
- Schräpler, J.-P.,** Benford's Law as an Instrument for Fraud Detection in Surveys Using the Data of the Socio-Economic Panel (SOEP), in: Jahrbücher für Nationalökonomie und Statistik, 231. Jg., Nr. 5-6, S. 212-234.
- Skousen, Christopher J./ Guan, Liming/ Wetzel, T. Sterling,** Anomalies and Unusual Patterns in Reported Earnings: Japanese Managers Round Earnings, in: Journal of International Financial Management and Accounting, 2004, S. 212-234.
- Sosna, Christian,** Statistische Ziffernanalyse -Teil I-, in: Steuerliche Betriebsprüfung, 09/2004, S. 249-253.
- Stephens, M. A.,** The goodness-of-fit statistic, VN: Distribution and significance points, in: Biometrika, 1965, Vol. 52, S. 309–21.
- Thomas, Jacob K.,** Unusual Patterns in Reported Earnings, in: The Accounting Review, 1989, S. 773-787.
- Tödter, Karl-Heinz,** Das Benford-Gesetz und die Anfangsziffern von Aktienkursen, in: WiSt, 2007, S. 93-97.
- Tödter, Karl-Heinz,** Benford's Law as an Indicator of Fraud in Economics, in: German Economic Review, 2009, S. 339-351.
- Trede, Marc/Watrin, Christoph/ Ullmann, Robert,** Ziffernanalyse und Chi-Quadrat-Anpassungstest in der steuerlichen Anwendung, in: Die Betriebswirtschaft, 2009, S. 701-716.
- Van Caneghem, Tom,** Earnings Management induced by cognitive reference points, in: British Accounting Review, 2002, S. 167-178.
- Van Caneghem, Tom,** The Impact of Audit Quality on Earnings Rounding-up Behaviour: Some UK Evidence, in: European Accounting Review, 2004, S. 771-786.
- Verband Deutscher Ingenieure,** VDI-Richtlinie 3633, Blatt 1, 1993.
- Von Sybel, Ludwig/ Roscher, Wilhelm Heinrich (Hrsg.),** Ausführliches Lexikon der griechischen und römischen Mythologie, Band 1,1, Leipzig 1886, S. 886 f.

- Watrin, Cristoph/Struffert, Ralf**, Benford's Law und Chi-Quadrat-Test – Chancen und Risiken des Einsatzes bei steuerlichen Prüfungen – Zugleich Anmerkungen zum Urteil des FG Münster vom 07.12.2005 1 K 6384/03 E, in: Der Betrieb, 2006, S. 1748-1750.
- Watrin, Christoph/ Struffert, Ralf/ Ullmann, Robert**, Benford's Law: an instrument for selecting tax audit targets?, in: Review of Managerial Science, 2008, S. 219-237.
- Watrin, Christoph/ Ullmann, Robert**, Ziffernanalyse in der steuerlichen Betriebsprüfung: Voraussetzungen, Funktionsweise, Anwendungsmöglichkeiten, in: Die Wirtschaftsprüfung, 2009, S. 98-106.
- Wähnert, Andreas**, Anwendbarkeit, Aussagekraft und Grenzen stochastischer Manipulationstests, in: Steuerliche Betriebsprüfung, 11/2008, S. 312-318.
- Wolf, Klaus**, Monte-Carlo-Simulation – Einsatz im Rahmen der Unternehmensplanung, in: Controlling 21/2009, S. 545-552.

Rechtsprechungsverzeichnis

BGH	14.12.1989	4 StR 419/89	Neue Juristische Wochenschrift, 1990, S. 1549
FG Münster	05.12.2002	8 V 5774/02	EFG 2004, 236
FG Münster	14.08.2003	8 V 2651/03 E,U	openJur 2011, 26048
FG Münster	10.11.2003	6 V 4562/03 E, U	EFG 2004, 236

Quellenverzeichnis

Autobild.de, Werkstattrechnung – Das Kürzel-Chaos, <http://www.autobild.de/artikel/werkstattrechnung-38427.html> (Abgerufen am 02.02.2016), 2002.

Bundeszentrale für politische Bildung, Der Weg in den Krieg, <http://www.bpb.de/geschichte/deutsche-geschichte/der-zweite-weltkrieg/199397/der-weg-in-den-krieg> (Abgerufen am 04.01.2015), 2015.

Bundeszentrale für politische Bildung, Die wirtschaftliche Entwicklung der Bundesrepublik Deutschland, <http://www.bpb.de/politik/grundfragen/deutsche-verhaeltnisse-eine-sozialkunde/138634/wirtschaftliche-entwicklung?p=all> (Abgerufen am 22.12.2015), 2012.

Bundeszentrale für politische Bildung, Massenmord und Holocaust, <http://www.bpb.de/izpb/151942/massenmord-und-holocaust> (Abgerufen am 09.02.2015), 2012.

Bundeszentrale für politische Bildung, Wer trägt die Schuld? – Schießbefehl und Mauertote, <http://www.bpb.de/geschichte/deutsche-geschichte/kontraste/42470/schiessbefehl-und-mauertote> (Abgerufen am 09.02.2016), 2012.

Bundeszentrale für politische Bildung, Versailles und Weimar, <http://www.bpb.de/apuz/30789/versailles-und-weimar?p=all> (abgerufen am 23.12.2015), 2008.

Bundeszentrale für politische Bildung, Wirtschaft und Gesellschaft unterm Hakenkreuz, <http://www.bpb.de/geschichte/nationalsozialismus/dossier-nationalsozialismus/39551/wirtschaft-und-gesellschaft?p=all> (Abgerufen am 28.03.2015), 2005.

Bundeszentrale für politische Bildung, Zerstörung der Demokratie 1930 – 1932, <http://www.bpb.de/geschichte/nationalsozialismus/dossier-nationalsozialismus/39537/zerstoerung-der-demokratie> (Abgerufen am 04.01.2016), 2011.

Die Bundesregierung, Die wirtschaftliche Lage der DDR vor der Wende 1989, <http://www.bundesregierung.de/Content/DE/Magazine/MagazinWirtschaftFinanzen/072/s1-die-wirtschaftliche-lage-der-ddr-vor-der-wende.html> (Abgerufen am 28.03.2015), 2009.

Distance Learning Center, Appendix – Table 7: Kolmogorov-Smirnov Test, <http://dlc.eri.eri.com/online/textbook/index.cfm?fuseaction=textbook.appendix&FileName=Table7> (Abgerufen am 16.03.2015), 2014.

Focus Money Online, Weniger Selbstanzeigen von Steuerbetrügnern - Druck soll bleiben (Abgerufen am 09.02.2016), 2015.

Focus Online, Streichen mit System, http://www.focus.de/finanzen/versicherungen/kfz-versicherung/recht-streichen-mit-system_aid_220818.html (Abgerufen am 29.03.2015), 2007.

Gesamtverband der Deutschen Versicherungswirtschaft, Scherpunkt Versicherungsbetrug, <http://www.gdv.de/versicherungsbetrug> (abgerufen am 21.01.2015), 2014.

Gesamtverband der Deutschen Versicherungswirtschaft, Versicherungsbetrug in der Kfz-Versicherung: Nicht jeder Unfall ist Zufall – Provozierte Verkehrsunfälle, <http://www.gdv.de/2008/04/versicherungsbetrug-in-der-kfz-versicherung-nicht-jeder-unfall-ist-zufall-provozierte-verkehrsunfaelle> (Abgerufen am 28.03.2015), 2008.

Handelsblatt-Online, Deutliche Preisunterschiede zwischen Markenwerkstatt und freiem Betrieb, <http://www.handelsblatt.com/auto/test-technik/autowerkstaetten-deutliche-preisunterschiede-zwischen-markenwerkstatt-und-freiem-betrieb/3052936.html> (Abgerufen am 29.01.2016), 2008.

haufe.de, Rekordjahr für Selbstanzeigen, http://www.haufe.de/steuern/kanzlei-co/steuerhinterziehung-rekordjahr-fuer-selbstanzeigen_170_287304.html (Abgerufen am 13.04.2015), 2015.

KPMG, Wirtschaftskriminalität in Deutschland 2014, <http://www.kpmg.com/de/de/bibliothek/presse/seiten/kpmg-studie-wirtschaftskriminalitaet-in-deutschland-2014.aspx> (abgerufen am 21.01.2015), 2014.

Manager Magazin, Beitrittsbetrug – Griechenland kommt ohne Strafe davon, <http://www.manager-magazin.de/unternehmen/artikel/a-328064.html> (abgerufen am 21.01.2015), 2004.

Mitteldeutscher Rundfunk, Blockparteien und Einheitslisten – Wahlen in der DDR, http://www.mdr.de/damals/eure-geschichte/themen/wahlen106_page-0_zc-6615e895.html (abgerufen am 28.03.2015), 2012.

Mitteldeutscher Rundfunk, Wie pleite war die DDR?, <http://www.mdr.de/damals/ddrpleite104.html> (Abgerufen am 28.03.2015), 2011.

Möller, Manuela, Measuring the quality of auditing services with the help of Benford's Law?, <http://ssrn.com/abstract=1529307> (Abruf vom 23.01.2015), 2009.

Morrow, John, Benford's Law, Families of Distributions and a Test Basis, <http://johnmorrow.info/pdf/benford.pdf> (Abgerufen am 17.12.2014), 2014, S. 4.

Posch, Peter N., Ziffernanalyse in der Fälschungsaufspürung. Benford's Gesetz und Steuererklärungen in Theorie und Praxis, http://posch.org/paper/posch_faelschungenbenford.pdf (Abfrage vom 28.03.2015), S. 16-34.

Spiegel-Online, Hyperinflation 1923 - Als die Mark vernichtet wurde, <http://www.spiegel.de/einestages/hyperinflation-1923-a-948427.html> (Abgerufen am 26.12.2015), 2009.

Süddeutsche Zeitung, Zwangsarbeit im Dritten Reich, <http://www.sueddeutsche.de/politik/deutschland-ein-sklavenhalterstaat-zwangsarbeit-im-dritten-reich-1.526445> (Abgerufen am 28.03.2015), 2008.

Eidesstattliche Erklärung

Ich versichere, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Weitere Personen waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder anderer Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer Prüfungsbehörde vorgelegt.

Ich bin darauf hingewiesen worden, dass die Unrichtigkeit der vorstehenden Erklärung als Täuschungsversuch bewertet wird und gemäß § 7 Abs. 10 der Promotionsordnung den Abbruch des Promotionsverfahrens zur Folge hat.

Ilmenau, der 28.02.2016

Daniel Brinkert